UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DEPARTMENT OF INFORMATION ENGINEERING

Master of Science in
Communication Technologies and Multimedia

CORRELATING CINEMATOGRAPHIC CAMERA TECHNIQUES AND ELICITED EMOTIONAL RESPONSE ON HOLLYWOOD MOVIE SCENES

Supervisor: Dr. Mattia Savardi
Supervisor: Dr. Sergio Benini

Candidate:
Aine van den Aakster
Matriculation Number: 744081

Academic Year 2024/2025

Correlating cinematographic camera techniques and elicited emotional response on Hollywood movie scenes

by
Aine van den Aakster

Submitted to the Department of Information Engineering
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN
COMMUNICATION TECHNOLOGIES AND MULTIMEDIA
at the
UNIVERSITY OF BRESCIA
March 2025

© 2025 Aine van den Aakster. All rights reserved.

The author hereby grants to the University of Brescia a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by:	Aine van den Aakster
Certified by:	Dr. Mattia Savardi
Accepted by:	Dr. Sergio Benini

Abstract

Through every decision of the creative process of filmmaking, directors carefully craft an audiovisual entertainment experience that evokes a range of emotional responses in its viewers. Ongoing research on the use of formal film features and their role in eliciting empathy in audiences supports filmmakers' decision-making and control in the emotional journeys they create through their medium. The use of the video camera in filmmaking determines the viewer's perspective on a scene through features like camera angle, camera level, shot scale, camera motion, etc. This can influence the viewer's comfort level, understanding of narrative context, perceived sense of intimacy or power dynamics between themselves and/or characters, and more. Data-collection of the current use of formal film features in mainstream cinema for this type of research has recently shifted towards various machine-learning approaches. The research aim is to establish a correlation between the use of Camera Angle and Camera Level as stylistic elements in contemporary, popular, empathy-eliciting Hollywood films by empirically describing their respective and combined uses to uncover stylistic conventions. Data on Camera Angle and Camera Level was automatically extracted from 22,454 individual film frames using a Convolutional Neural Network (CNN) model. By implementing a custom data augmentation pipeline, the training dataset was able to fine-tune a pre-existing CNN model, achieving satisfactory validation accuracy, precision, recall, and accuracy. The extracted data was then analyzed using both descriptive and correlational statistical methods. Findings show that half (50.80%) of all film frames are captured at a Neutral Angle, a strong dominance over other angle classes. In terms of Camera Level, Shoulder Level (38.63%) and Eye Level (33.61%) are the most prevalent. Conversely, Overhead Angle (2.29%) and Ground Level (2.85%) are notably uncommon. When examining Camera Angle and Camera Level as combined pairings, results indicate that the most frequent classifications often appear together. Overall, statistical analysis reveals a moderate yet highly significant correlation between Camera Angle and Camera Level, with a Chi-Squared test p-value of approximately 0 and a Cramér's V of 0.31. Clear statistical patterns emerge in the pairing of level-angle throughout the data. More neutral Level classes (Eye, Shoulder and Hip Level) are predominantly matched with High, Neutral, or Low Angles. Likewise, more extreme classifications of either feature (such as Aerial, Knee, and Ground Level, or Overhead and

Dutch Angle) tend to be paired with each other. Additional pairing conventions align naturally with the framing of human subjects. The results of this study contribute to a broader understanding of how formal film techniques are employed in contemporary, popular, empathy-eliciting Hollywood films. We anticipate this research to be starting point for more elaborate studies correlating formal camera features to emotional data collected from film viewers. This, in the future, could lead to accurate predictions of audience's emotional reception of films, a potentially powerful tool in the pre-production, production and post-production stages of filmmaking, as well as in other disciplines related to video content.

Contents

Introduction **1**

1 Film analysis and emotional state **5**
- 1.1 Formal Features of Film 5
- 1.2 Camera Angle 6
- 1.3 Camera Level 7
- 1.4 Formal Feature Extraction 7
- 1.5 State of the Art 8

2 Data **11**
- 2.1 Pre-existing datasets 11
- 2.2 Composition of the dataset 12
- 2.3 Annotation of the dataset 12
- 2.4 Annotation criteria 14
  - 2.4.1 Camera Angle 14
  - 2.4.2 Camera Level 15
  - 2.4.3 Ambiguous cases 15
- 2.5 Dataset supplement 16
- 2.6 Data Augmentation 16

3 Methods **19**
- 3.1 Camera Angle and Camera Level extraction 19
  - 3.1.1 Pre-trained CNN Model 19
  - 3.1.2 Initial fine-tuning of the model 19
  - 3.1.3 Final fine-tuning of the model 20
- 3.2 Statistical Analyses 20
  - 3.2.1 Descriptive Statistics 20
  - 3.2.2 Correlational Statistics 20

4 Results **23**
4.1 Model fine-tuning 23
4.1.1 Initial fine-tuning of the model 23
4.1.2 Final fine-tuning of the model 23
4.1.3 Error analysis 28
4.2 Statistical analysis 28
4.2.1 Descriptive statistics 28
4.2.2 Correlational statistics 29

Conclusions and future work **35**

A Initial fine-tuning **37**

B Final Fine-tuning **49**

C Temporal visualization of Camera Angle/Camera Level pairing **61**

Introduction

> “ The great gift of human beings is
> that we have the power of empathy,
> we can all sense a mysterious
> connection to each other. ”
>
> *Meryl Streep,*
> *Hollywood actress*

Directing Emotional Journeys

The movie industry is indisputably one of the biggest when it comes to entertainment industries. Its global box office revenue, expected to reach $33 billion in 2025, has been steadily climbing back up to pre-COVID-19 figures [1]. There is certainly money to be made in leaving people feeling entertained and satisfied. The entertainment value of cinema works on multiple sensory and cognitive levels. Through carefully orchestrated visual and auditory cues, cinema can trigger profound emotional arousal. It is precisely this emotional evocation that is one of the most obvious types of gratification cinema provides.

The emotional responses evoked in cinema audiences throughout a film's narrative are founded on empathy. It is generally understood in neuropsychology that empathy is a core affective-cognitive process through which another's mental state can be interpreted, like that of an on-screen character. [2][3]. Cognitive empathy relates to imagination and deliberate perspective-taking without the need for shared affect. This is a crucial and thus highly refined skill for humans, which allows audiences to feel so deeply for fictional scenarios presented to them[4]. The range of emotions evoked in this way through cinema varies greatly across a plethora of cinematographic genres. Audiences even voluntarily seek out unpleasant emotions, such as fear, sadness or disgust, paradoxically still experiencing pleasure from the movie-watching activity as a whole. This is because the emotional experiences of viewers can be rewarding in multiple ways, such as through fun, thrill, catharsis or a subconscious desire to practice certain social skills [5][6][7].

Rewarding experiences of any nature lead to positive content evaluation, which in turn leads to success and revenue for the film [6]. In order to provide such experiences to its viewers, popular cinema typically incorporates empathy-eliciting techniques. Empathic engagement from viewers is thought to foster investment in the film's unfolding narrative [8]. It enhances comprehension of the plot of the film by facilitating interpretations of film characters' mental states [9], acts as a mediator between narrative exposure and media effects [10] and as a predictor of narrative persuasion effectiveness [11]. It is therefore a valuable skill as cinematographer to know how to incorporate empathy-eliciting techniques effectively through the medium of cinema, in order to maximally entice an audience to step into the film's world. This skill entails aiming the creative decisions involved in mainstream filmmaking towards the goal of eliciting empathy, effectively directing the viewer's emotional experience [12]. Expanding knowledge on these creative decisions through the lens of empathy elicitation may thus contribute to create successful films that meet entertainment goals and revenue targets.

Creative decisions are present in any and all facets of the filmmaking process, from script-writing to post-production, many of which are shared with other media such as literature or music. In filmmaking specifically, through the use of a video camera, the interaction of story and writing with the unique audiovisual presentation of the cinematic medium plays an important role in evoking emotions [13]. As such, formal features of film, defined by Lankhuizen et al. as 'the stylistic and aesthetic properties of the presentational format, such as cinematography and editing' [14], are critical in shaping emotional audience responses. Some of the roles played by certain formal features like shot scale and shot duration [15][16][17], lighting, color and motion [18] were researched in the last two decades, though relatively few studies directly relate formal features to empathy-elicitation in popular fiction films.

The Aim of this Research

This research serves to contribute to a general understanding of how formal features are currently used in mainstream empathy-eliciting cinema. Specifically, it is the aim of this research to investigate the use of Camera Angle and Camera Level, in contemporary, popular, empathy-eliciting Hollywood (CPEEH) films. Empirically describing the use of these two formal camera features throughout CPEEH film scenes may reveal stylistic conventions that characterize this type of cinema. This is done through an analysis of descriptive statistics as well as correlational statistics for both features. The statistical findings may contribute to an increased understanding of stylistic choices in CPEEH cinematography, supporting better informed choices in the creative processes of filmmakers. This may open up possibilities for future developments that facilitate control and precision in directing emotional journeys.

Thesis outline

This paper presents the findings of this research on the use of Camera Angle and Camera Level in CPEEH films. Chapter 1 presents a theoretical background on the principal concepts related to this study. In particular, Camera Angle and Camera Level and their relation to emotional viewer responses. Through this, some important definitions are established for this paper. It then outlines the motivation and objectives of video processing, followed by a state of the art analysis for this research topic. In Chapter 2 the data used in this study is described. Chapter 3 presents the methodology of data collection, processing and analysis as well as the implementation or use of any tools used in these phases. The results for each method are presented in Chapter 4. Finally, the last chapter, Conclusions and Future Work, reiterates the main results before discussing the limitations of this research and suggestions for future research.

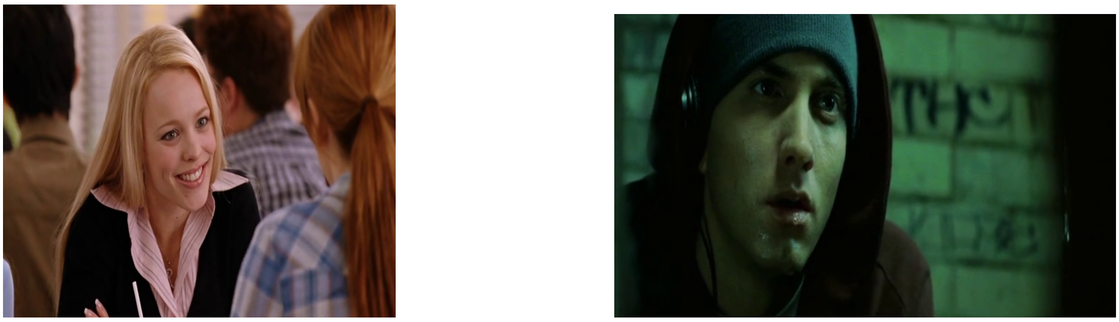
Chapter 1

Film analysis and emotional state

In this chapter, a theoretical background is presented on the principal concepts of this research, establishing how these concepts are defined in this research where necessary. In particular, the relevance of Camera Angle and Camera Level to eliciting empathy is argued. Then, the motivation and objectives of video processing are outlined. Finally, a state of the art analysis of related research is presented.

1.1 Formal Features of Film

As a general definition, formal features are the noncontent part of media. Whereas content refers to the story, characters, plot and actions, formal features further shape a film through aural and visual elements and structures [19]. This paper adopts the Lankhuizen et al. definition of formal features as *'the stylistic and aesthetic properties of the presentational format'*. In film, this entails anything from camera techniques to musical scores, from lighting to pacing and edits [20]. This large quantity of features can be combined into an abundance of different ways, allowing a wide variety in the overall look-and-feel of films. A side-by-side comparison of two stylistically differing film scenes is shown in Figure 1.1.



(a) Mean Girls (2004)

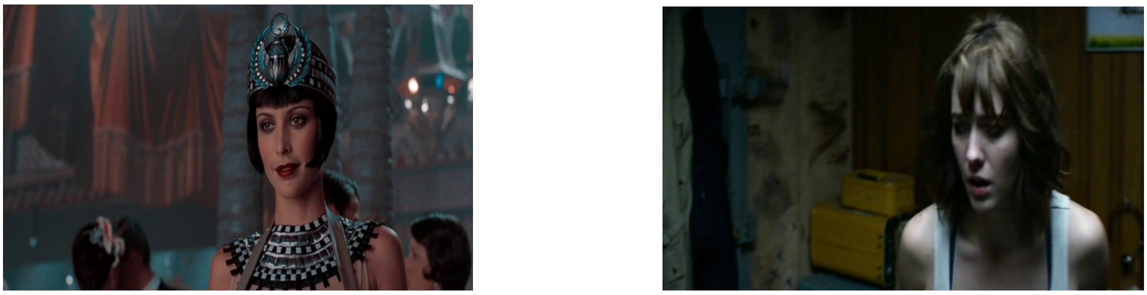
(b) 8 Mile (2002)

Figure 1.1: A side-by-side comparison of two stylistically differing CPEEH film shots.

These formal features are then further divided into *high-level features* and *low-level features*. High-level features relate to structure and semantics whereas low-level features relate to craft and technique. Camera Angle and Camera Level, which are the focus of this research, are both low-level visual features, directly related to the use of the camera while shooting the film. As will be described in sections 1.2 and 1.3, Camera Angle and Camera Level may have an emotional impact on the audience. Cinematographers make deliberate use of these effects [12]. More knowledge of the impact on viewers may directly benefit their work but might also benefit other disciplines within the film industry or otherwise related to film and empathy. Future applications might be found in film audience suitability rating organizations, advertising or various kinds of visual media content creation.

1.2 Camera Angle

Camera Angle, in this research, is defined as the rotation of the camera along the lateral or longitudinal axis. The latter, a tilt to the right or the left, is referred to as a Dutch Angle, after its roots in silent-era expressionist German ('Deutsch') films [21]. With time, 'Deutsch' turned into 'Dutch'. Viewers will experience uneasiness and disorientation. This tilt is used as a metaphor for any kind of instability present in the narrative. It can be used when a character is drugged or sick, or something about the scene is "not quite right", keeping audiences on edge [22]. Studies have indicated that the Camera Angle along the lateral axis can affect character evaluation. Low-angle shots —where the viewer looks up at a character— can make characters appear more dominant, stronger, bolder, more aggressive, and less fearful. In contrast, high-angle shots, which position the viewer above the character, imply weakness and vulnerability [23][24]. A side-by-side comparison of a High Angle and a Low Angle CPEEH film shot can be found in Figure 1.2.



(a) The Aviator (2004)

(b) 10 Cloverfield Lane (2016)

Figure 1.2: A side-by-side comparison of a High Angle and a Low Angle CPEEH film shot.

In this way, Camera Angle establishes a power dynamic between characters or between the viewer and a character. Lastly, a neutral Camera Angle can create a sense of parity, which may foster empathy through a sense of similarity between the viewer and the character [23]. This suggests that Camera Angle plays a role in shaping audiences' empathic engagement by altering perceptions of a character or situation. An overview containing examples of Camera Angle shots is shown in Figure 1.3.

1.3 Camera Level

Camera Level establishes the viewer's perspective on a scene. In this paper, it is defined as the height at which the camera is positioned relative to the ground. Eye Level and Shoulder Level are considered neutral, typically used to depict natural conversations between characters. These levels tend to place the viewer among the characters and as a consequence the viewer will relate to the characters as equals [23]. Hip level shots can be quite a comfortable and natural perspective on non-standing characters, but can also be used to emphasize something of visual or dramatic interest at this height. Knee or Ground level shots often offer a limited view of a scene, inviting the viewer to imagine things out of view, such as a character's face. In contrast to Ground Level, Aerial level shots usually contain a large part of the surroundings, offering the viewer a sense of space, time or reality. The use of Camera Level can affect the empathic response of a viewer, by visually underscoring the story. An overview containing examples of Camera Level shots is shown in Figure 1.3.

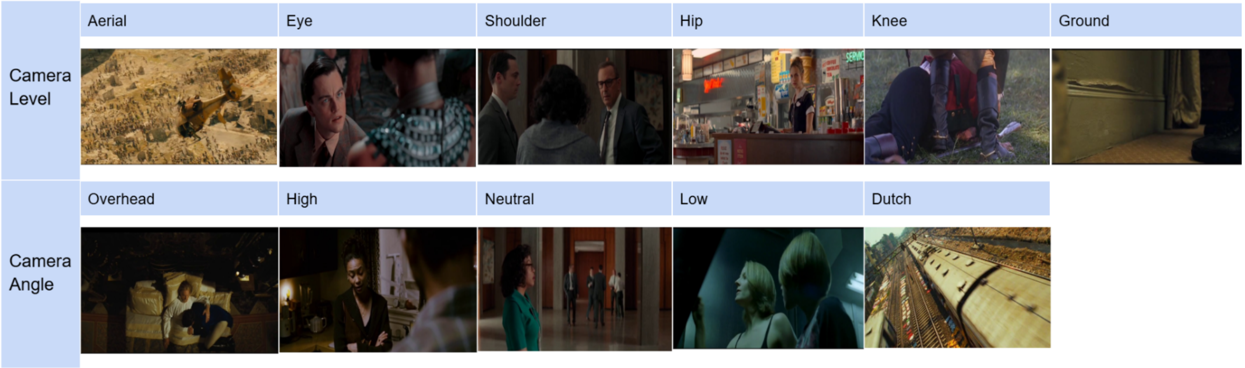


Figure 1.3: Sample shots from each class of Camera Angle and Camera Level.

1.4 Formal Feature Extraction

In order to conduct empirical research on the use of formal features in film, quantitative and/or qualitative data needs to be collected. This can be done by manually annotating films, scenes, shots or even individual frames. However, this approach comes with limitations in time and resources. These limitations can be mitigated by crowd sourcing the annotations [25], though this introduces new challenges of quality control. Recent developments in the field of machine learning have opened up possibilities for automated video annotation. A growing body of research is proposing and evaluating ways to implement machine learning for the automated collection of video data such as scene segmentation [26], soundtrack suitability [27], genre classification [28] or shot scale extraction [29]. This research implements the Convolutional Neural Network (CNN) recently proposed by Savardi et al. for the task of automated Camera Angle and Camera Level extraction of single film frames [30].

1.5 State of the Art

Cinemetrics is a term originally referring to the collection of movie data, but which over the past two decades has shifted to include the visualization of movie data, now regarded as a data art form [31] [32] [33]. An example of a cinemetrics project is Frederic Brodbeck's 2011 *Cinemetrics* [34]. Brodbeck describes its aim as *'measuring and visualizing movie data, in order to reveal the characteristics of films and to create a visual "fingerprint" for them'*. A sample visualization of *Cinemetrics* is shown in Figure 1.4. Other notable projects in this field are *Cinema Redux* [35], *MOVIEBARCODE* [36] and *The Colors of Motion* [37], though these are all completely based on movies' color palettes and lacking in empirical analyses. Visualizations of the aforementioned projects are shown in Figure 1.5. Empirical research on camera technique based formal features in relation to empathy remains limited. A notable exception to this is the effect of Camera Angle on object evaluation, which has been well-researched for decades, for marketing purposes [38] [39] [40]. In recent years, however, Machine Learning is proving to be a valuable approach for the data collection of formal camera features, giving this field of research a boost. In 2020, a paper was

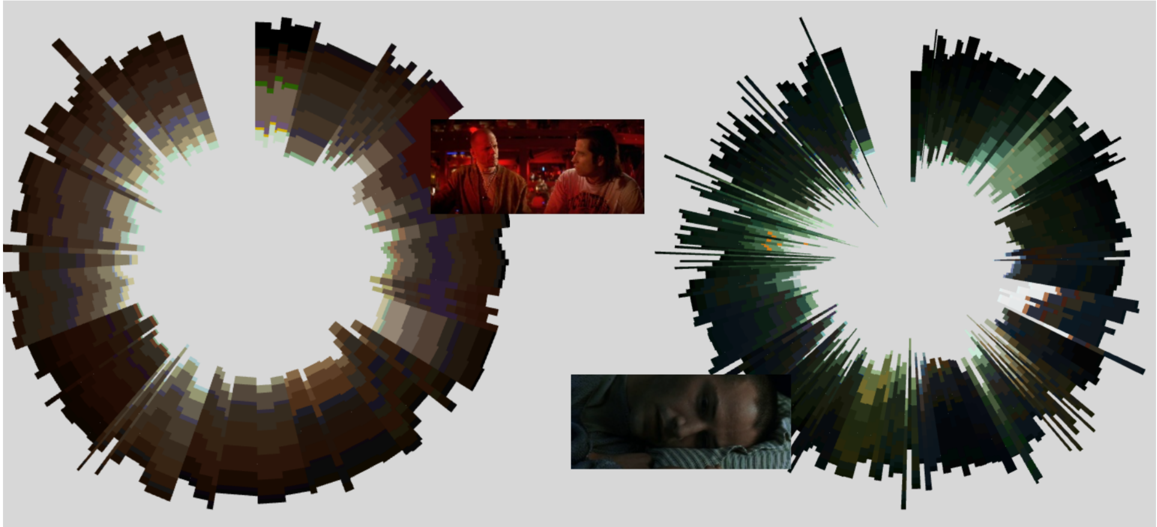
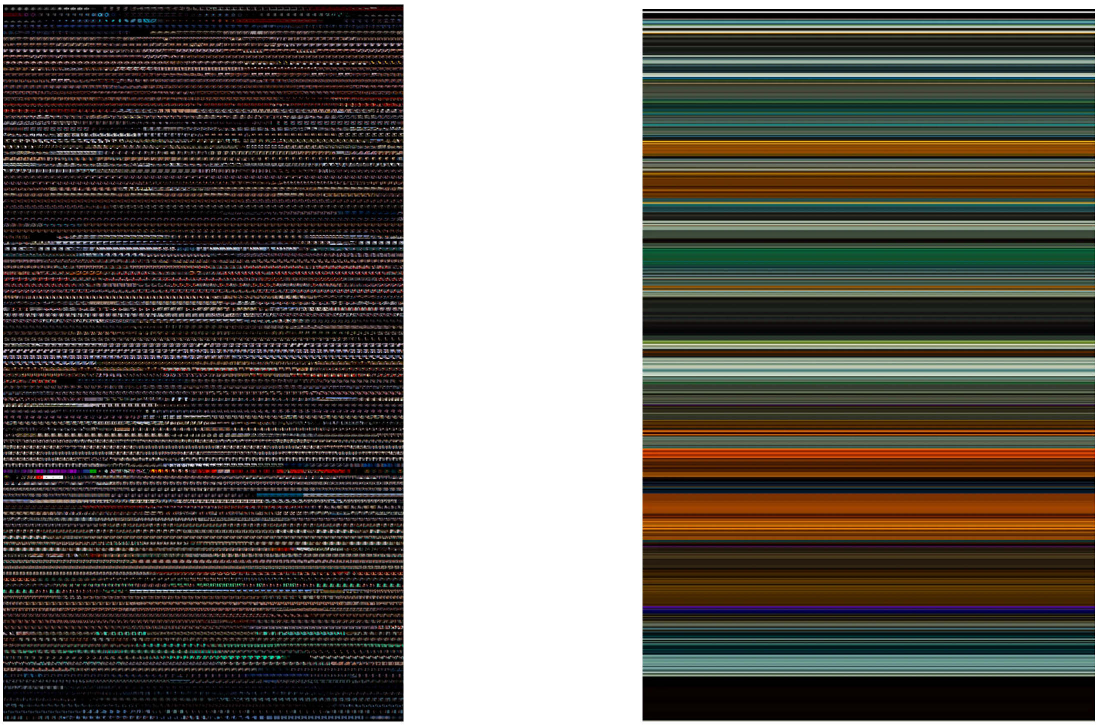


Figure 1.4: *Cinemetrics* visualizations of Pulp Fiction (1994) and The Matrix (1999)

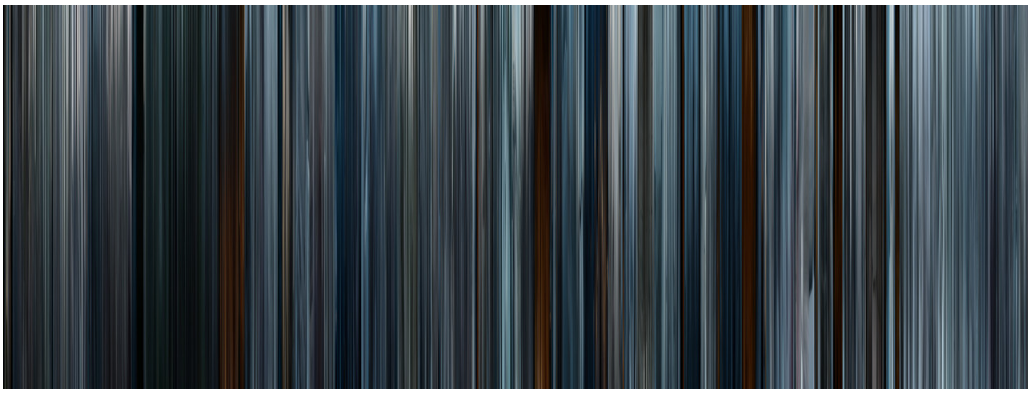
published on shot type recognition using a learning framework Subject Guidance Network (SGNet) [41]. SGNet separates the background from the subject of a shot as two separate guidance maps, allowing accurate movement type and scale classification. A 3D CNN was proposed in 2021 for the automatic classification of camera motion (travel, pan, roll, zoom, etc.) [42]. The proposed model showed high performance and robustness, with an average accuracy rate of 94%. These machine-learning approaches are supporting an expansion in the research of formal film features relating to empathy. With the proposal of an automated shot scale extraction framework in 2016 [43], research on the emotional effects of shot scale has expanded. Its influence has now been established on viewers' theory of mind responses, mood and narrative engagement [44] [15]. A broad, exploratory quantitative formal analysis of CPEEH film scenes was conducted in 2022. That analysis considered various camera-based formal features such as Shot Scale, Camera Angle and Camera Perspective, though the latter two had been coded manually [14]. Since then, a 2023 paper has proposed automated extraction of Camera Angle as well as Camera Level with the use of a CNN [30]. The data used to train that model was largely based on mainstream fiction film, which makes it a suitable starting point for automated Camera Angle and Camera Level extraction on the data of this research.

The datasets of the aforementioned exploratory analysis as well as that of the paper on automated Camera Angle and Camera Level extraction are used in this research. The next chapter describes the nature of these datasets, outlining how the data was originally obtained, and how it was combined and augmented for this research.



(a) *Cinema Redux* visualization of Vertigo (1958)

(b) *The Colors of Motion* visualization of Blade Runner 2049 (2017)



(c) *MOVIEBARCODE* visualization of The Revenant (2015)

Figure 1.5: Visualizations of various cinemetrics projects.

Chapter 2

Data

This Chapter describes the data used in this research, originating from two different pre-existing datasets. The first dataset was taken from an exploratory formal analysis of 100 CPEEH film scenes. The second dataset was taken from the research proposing the CNN model applied in this research. The following sections outline how these respective datasets were obtained by the original researchers and how the data was combined, annotated and augmented for this research.

2.1 Pre-existing datasets

The data for this research was mainly obtained from a dataset containing 100 film scenes from 100 distinct CPEEH films, from Lankhuizen et al. [14]. In the paper, Lankhuizen et al. describe how 100 films were randomly selected from a pool of films that met certain criteria to be considered contemporary, popular, empathy-eliciting and Hollywood cinema. These films were then segmented temporally into ‘scenes’, referred to as discrete narrative units. These scenes were then again filtered by certain criteria, among others, being a coherent and independent story unit, containing sufficient plot-relevant information and having at least one identifiable main character whose state of mind is communicated to the audience in some way. Finally, a random selection of one of these scenes from each of the 100 films, resulted in a 100 empathy-eliciting film scene dataset used in this research.

Images were also taken from the annotated Ground Truth training set of images used to train the pre-existing CNN [30]. All frames from classes Eye Level, Shoulder Level and Hip Level were automatically sampled from the CineScale database [45]. Rarer classes, such as Ground Level and Knee Level, were populated by manually filtering Google Image Search results. Aerial Level shots were extracted from public online videos containing drone footage of various landscapes and cities. Other classes were populated by scraping the One Perfect Shot database of Film School Rejects [46].

2.2 Composition of the dataset

From all CPEEH film scenes [14], frames were extracted at one second intervals, resulting in a 22,454 frame dataset. A subset of this was annotated to be used as Ground Truth for training the model, Dataset ANN. Class imbalances were diminished by supplementing smaller classes with additional images from the aforementioned dataset of Savardi et al.[30], resulting in Dataset BAL. Later, based on a first round of model training performance metrics, the dataset was increased one last time through data augmentation. This final Dataset AUG then served as Ground Truth for the final fine-tuning of the model. A quantitative overview of the datasets used can be found in Table 2.1.

	Overhead	High	Neutral	Low	Dutch	Total
Dataset ANN	22	222	494	231	54	1023
Dataset BAL	200	222	494	231	200	1347
Dataset AUG	400	399	901	462	400	2562

(a) Camera Angle

	Aerial	Eye	Shoulder	Hip	Knee	Ground	Total
Dataset ANN	36	330	354	277	53	29	1079
Dataset BAL	200	330	354	277	200	200	1561
Dataset AUG	400	484	533	494	400	400	2711

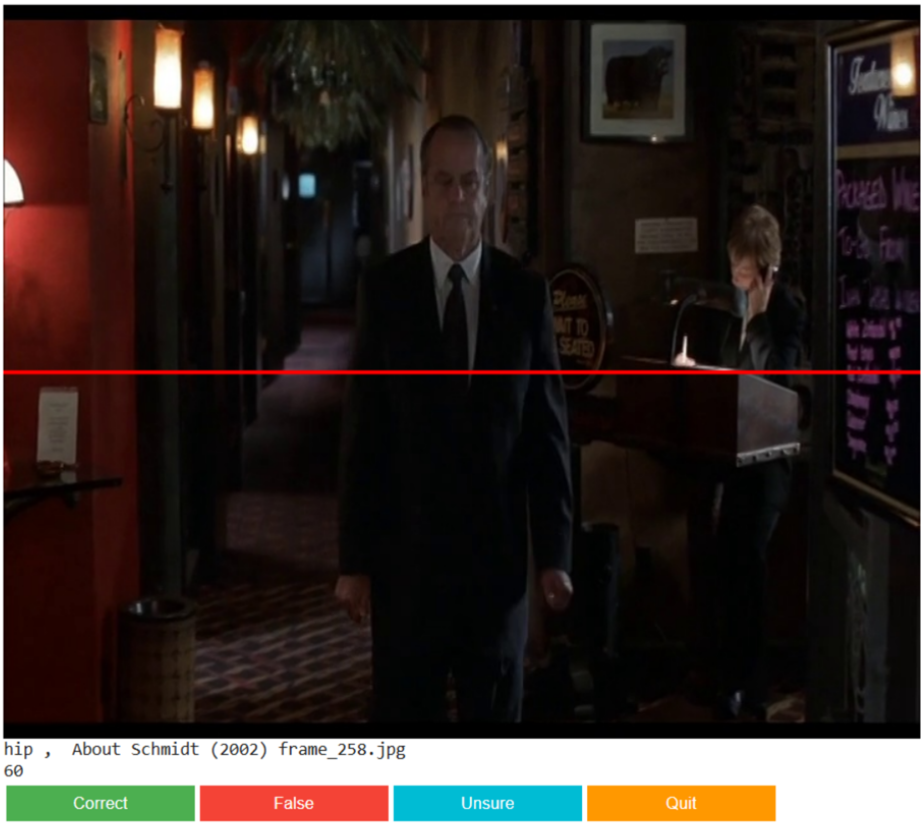
(b) Camera Level

Table 2.1: Quantitative overview of the image datasets created during this research.

2.3 Annotation of the dataset

Firstly, a pre-existing model[30] was applied to get initial classifications of Camera Angle and Camera Level for each frame. After a quick glance it was clear that performance was very poor, though for a lack of Ground Truth there was no quantitative analysis of this performance. From these original classifications, randomly selected frames from one class at a time were examined. Their annotations were corrected by hand where needed through a two-phase correction. As such, each annotation was considered at least at two different times before being finalized, and experienced researchers Benini and Savardi were consulted for ambiguous cases. In order to streamline the manual annotation process, interactive Python notebooks were developed for each phase. Any datasets referred to in this section are all of the same basic structure, mimicking that of Savardi et al.. They all have two main directories for Camera Angle and Camera Level, each containing subdirectories corresponding to their respective class labels. The dataset corresponding to the initial classification will be referred to as Dataset INIT. The Python notebooks were used to display random images from a specified directory one image at a time, alongside buttons to either copy or move the image file to other directories, depending on the phase. A horizontal red guideline was displayed across the middle of the images to aid in classification.

In the first phase, three additional subdirectories were created in each of the class label directories of the manually annotated dataset (Dataset ANN): ‘correct’, ‘incorrect’ and ‘unsure’. Images from a specified class label directory of Dataset INIT were examined and always copied to the Dataset ANN directory corresponding to the same label, but inside were split into three subfolders: ‘correct’, ‘incorrect’ and ‘unsure’. A screenshot of the interactive Phase One notebook is shown in Figure 2.1. During this phase, a set of criteria for the classification of each class was defined and agreed upon with Benini and Savardi.

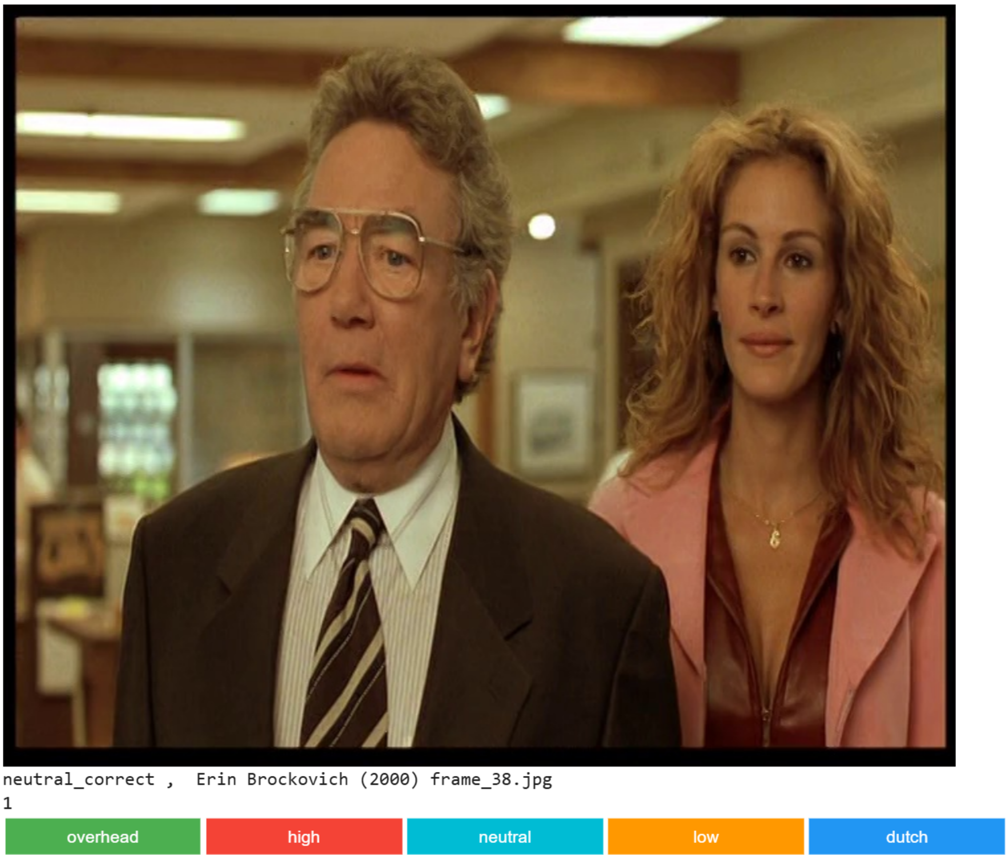


```
hip ,  About Schmidt (2002) frame_258.jpg
60
```

Figure 2.1: Phase One notebook displaying images from the Dataset INIT ’hip’ directory.

In the second phase, images from Dataset ANN were examined one subdirectory at a time and, according to the agreed upon criteria, moved to a class label directory. As such, all images examined in Phase One were considered again in this phase before finalizing their classifications, including the images that had already been placed in ‘correct’ subdirectories. A screenshot of the interactive Phase Two notebook is shown in Figure 2.2.

Based on the quantitative distribution of the classes, Phase One and Phase Two were repeated focusing on underrepresented categories until a total count of roughly 1000 images was reached for both Camera Angle and Camera Level.



```
neutral_correct ,  Erin Brockovich (2000) frame_38.jpg
1
```

Figure 2.2: Phase Two notebook displaying images from the Dataset ANN 'neutral/correct' directory.

2.4 Annotation criteria

The following annotation criteria were implemented while building Dataset ANN.

2.4.1 Camera Angle

All but one Camera Angle classes describe a rotation of the camera along the lateral axis. Dutch instead describes a rotation along the longitudinal axis.

- Overhead: Shots where the camera is pointed downwards, almost perpendicular to the ground.
- High: Shots where the camera is tilted downwards towards the ground, on an angle ranging from slightly tilted to almost overhead.
- Neutral: Shots where the camera is tilted neither downwards nor upwards.
- Low: Shots where the camera is tilted upwards.
- Dutch: Shots where the camera is rotated along the longitudinal axis.

2.4.2 Camera Level

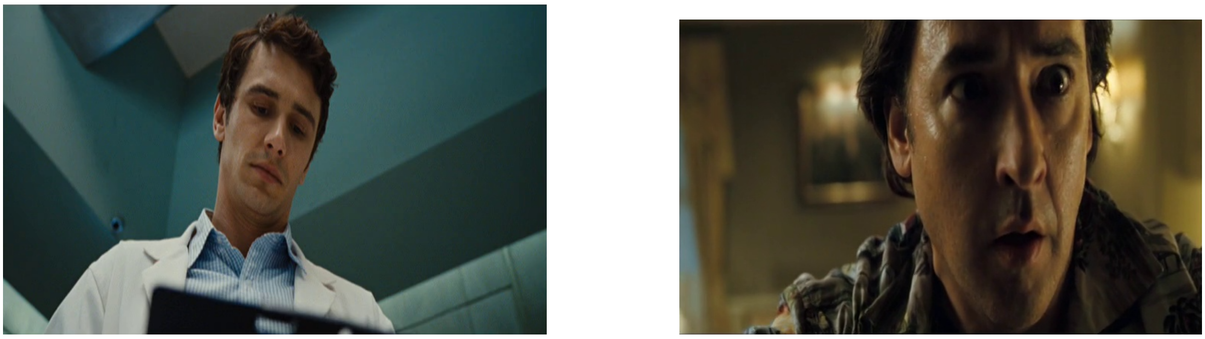
The various Camera Levels named in relation to body parts were considered in relation to the camera operator, presumed to be standing upright on the ground, and independently from any present subjects or the shot's Camera Angle.

- Aerial: Shots where the camera's vertical position is a considerable distance off the ground, for example from a drone or plane. These shots show a large part of the surroundings.
- Eye: Shots where the camera is at least eye level off the ground. This includes shots from heights that are clearly above eye level, but not quite high enough to be considered aerial shots.
- Shoulder, hip, knee: These are shots from shoulder, hip or knee height off the ground.
- Ground: Shots where the camera is close to the ground (closer than knee level).

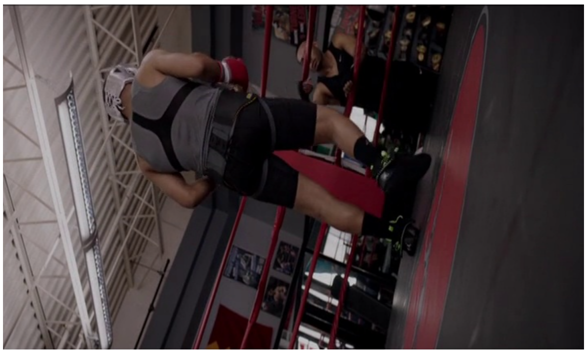
Since Camera Levels are not considered in relation to the subject, a high or low angle can cause the camera to be pointed towards a subject's body that differs from the Camera Level. For example, a low angle shot with a (standing) subject's eyes right on the middle guideline cannot be an eye level shot, instead the camera must have been at hip or shoulder height to be pointed upwards at the subject's eyes. Sample shots of some Camera Angle and Level pairings can be found in Figure 2.3.

2.4.3 Ambiguous cases

A camera's rotation along a certain axis or height off the ground is in reality continuous. This causes some edge cases during categorical annotation. For example, at what height does eye level become aerial, at what angle does an overhead angle become a high angle? Given the frequent nature of neutral angle eye level shots, it could be beneficial in the future to consider an additional transitionary class between eye level and aerial to describe shots that are not quite aerial, but where the camera's height is clearly above eye-level. Defining Camera Level as height off the ground raises questions concerning where the ground is. For example, what is considered ground level in shots of astronauts floating in space or when the scene contains slanted terrain or stairs? Some of these difficulties can be resolved by considering Camera Level in relation to the shot's subject. However, this raises other difficulties like having multiple, differently positioned subjects in a single shot (one sitting, one standing) and overlapping classes for any non-standing subjects. Furthermore, Camera Levels Eye and Shoulder are difficult to distinguish in long shots (shots where the camera is positioned at quite a distance from its subject(s)). Lastly, a



(a) Hip Level Low Angle shot (b) Eye Level Neutral Angle shot



(c) Ground Level Dutch Angle shot

Figure 2.3: Sample shots of various Camera Angle/Camera Level pairings.

Dutch angle corresponds to a rotation along the longitudinal axis, which can occur in combination with any rotation along the lateral axis, causing overlapping classes.

2.5 Dataset supplement

In order to diminish the class imbalances present in Dataset ANN, the weakest classes were supplemented with images from the Savardi et al. dataset [30]. Random images from corresponding classes were inserted into the dataset until each class contained at least 200 images, resulting in Dataset BAL.

2.6 Data Augmentation

To enhance generalizability and robustness of the model, data augmentation was applied to the dataset. Misclassifications of the testing phase of the best model from the initial fine-tuning were inspected to identify problematic classes, along with the testing classification report results. For this purpose, another interactive Python notebook was developed, displaying misclassified images in a grid where each cell corresponds to a specific Truth/Prediction pair. Inside each non empty grid cell, the corresponding misclassified images are displayed one at a time with a button to cycle through them. A screenshot of the interactive Misclassification notebook displaying a misclassification grid for Camera Angle is shown in Figure 2.4. Due to the similar nature and ambiguity between class pairs like eye-shoulder, shoulder-hip, high-neutral and neutral-low, some misclassification close to the diagonal is expected and acceptable in small quantities. Misclassifications further away from the diagonal have more potential to be strongly diminished or eliminated altogether as a result of data augmentation.

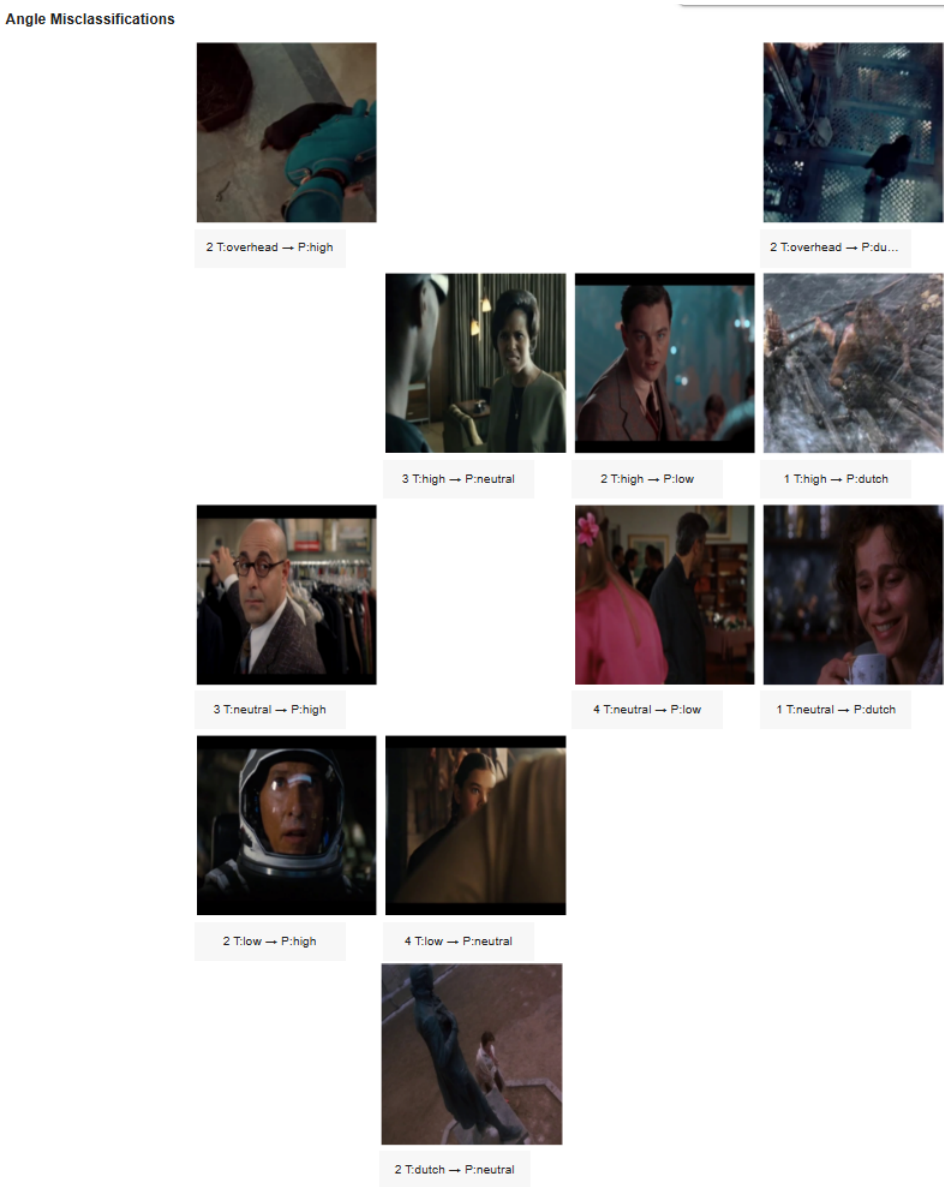

Angle Misclassifications
2 T:overhead → P:high
2 T:overhead → P:du...
3 T:high → P:neutral
2 T:high → P:low
1 T:high → P:dutch
3 T:neutral → P:high
4 T:neutral → P:low
1 T:neutral → P:dutch
2 T:low → P:high
4 T:low → P:neutral
2 T:dutch → P:neutral

Figure 2.4: Misclassification notebook displaying a Camera Angle misclassification grid.

A data augmentation pipeline was tailored according to the nature of the dataset being film frames, using image augmentation Python library *imgaug* [47]. In order to preserve the contextual integrity of different Camera Angles and Camera Levels, transformations like rotations or vertical flipping were avoided. Instead, the augmentation techniques applied were designed to mimic realistic spatial variations, different lighting conditions and movement-induced distortions. The configuration of the augmentation pipeline consisted of horizontal flipping (50% probability), brightness and contrast adjustments (50 percent probability, using random multiplication factors of 0.7-1.3 per channel), motion blur (50 percent probability) and hue and saturation adjustments (30 percent probability). This augmentation process was applied selectively, wherein classes with fewer samples and/or a lack of performance were augmented more strongly. An overview of the application strengths of the augmentation per class is shown in Table 2.2. This data augmentation process resulted in Dataset AUG.

Camera Angle	augmentation strength
Overhead	1.0
High	0.8
Neutral	0.8
Low	1.0
Dutch	1.0

(a) Camera Angle

Camera Level	augmentation strength
Aerial	1.0
Eye	0.5
Shoulder	0.5
Hip	0.8
Knee	1.0
Ground	1.0

(b) Camera Level

Table 2.2: Data augmentation application strengths per class.

The following chapter details the pre-existing CNN for Camera Angle and Camera Level extraction and describes the methodology of fine-tuning the model on the above presented data. It then outlines the subsequently performed statistical analyses of these formal features.

Chapter 3

Methods

In this Chapter, the methods applied in this research are detailed. Firstly, the fine-tuning of the pre-existing model is debribed. Secondly, the methods used for statistical analysis are explained for both descriptive statistics and correlational statistics.

3.1 Camera Angle and Camera Level extraction

3.1.1 Pre-trained CNN Model

The pre-existing model from Savardi et al. [30] is a CNN architecture with a ResNet18 backend. It handles the sets of classes for Camera Angle and Camera Level distinctly with two independent classification heads. It was originally trained with a custom loss function considering both classification head outputs in a weighted sum of their respective categorical cross entropies. During training iterations the two datasets were alternated (round-robin). This exploited their interdependence and ensured even training. A transformation was applied to all images in the dataset resizing them to 256x256 and centre-cropping them to 224x224 pixels. After the last convolutional layer of the ResNet18, its feature maps are passed on a fully connected layer (FC). This FC layer consists of 512 hidden units with ReLU activation. The final layer is an FC layer with softmax activation and respectively five and six neurons for the Angle head and the Level head.

3.1.2 Initial fine-tuning of the model

All fine-tuning of the model has been performed using Python library Pytorch [48]. Dataset BAL was initially used to fine-tune the pre-existing model. A transformation was applied to all images in the dataset resizing them to 256x256 and centre-cropping them to 224x224 pixels. Since the dataset is relatively small (<3000), a batch size of 64 was used. Each round of fine-tuning the dataset was split in 70/10/20 partitions for training, validation and testing respectively. Four separate rounds of training were executed. All executions employed the Adam optimizer, but with varying initial learning rates and numbers of epochs. Performances of each hyperparameter configuration can be found in Appendix A. The best fine-tuned model was determined based on its validation accuracy curve over its training epochs.

3.1.3 Final fine-tuning of the model

Restarting from the pre-existing (not fine-tuned) model, fine-tuning was performed again, this time on Dataset AUG, again with 4 different configurations of initial learning rate and maximum number of epochs. The best model was again determined by the validation curves. This model was thus applied to obtain Camera Angle and Camera Level predictions of all 22.454 frames for various statistical analyses.

3.2 Statistical Analyses

Finally, the Camera Angle and Camera Level data of the full 22.454 image dataset were analyzed in two parts. Firstly, to establish an initial understanding of the statistical properties of the dataset, descriptive statistics were calculated for Camera Angle and Camera Level separately, as well as scene duration. Then, the correlation between Camera Angle and Camera Level was investigated through various correlational statistical methods.

3.2.1 Descriptive Statistics

Standard descriptive statistics were calculated for scene duration, that is, mean, median and standard deviation as well as all quartile partitions. The distribution of scene durations was visualized with a histogram. The occurrence frequencies of the various classes of Camera Angle and Camera Level were computed. The most commonly used Camera Angles and Levels and the distribution of scene durations identify stylistic tendencies in CPEEH filmmaking.

3.2.2 Correlational Statistics

Co-occurence matrices were plotted to identify commonly occurring Camera Angle/Camera Level pairings. A Chi-Squared test of independence was conducted to determine whether there is a statistically significant relationship between Camera Angle and Camera Level in the dataset. This test assesses whether the distribution of one categorical variable differs depending on the levels of another [49]. If a significant association is found, it suggests that certain Camera Angles are more likely to appear with specific Camera Levels.

To quantify the strength of this association, Cramér's V was calculated. Cramér's V is a measure of effect size for categorical variables, ranging from 0 (no association) to 1 (perfect association) [50]. Principal component analysis (PCA) was performed on one-hot encoded predictions to identify underlying patterns in the dataset. PCA is a dimensionality reduction technique that transforms high-dimensional data into a smaller set of principal components while preserving as much variance as possible [51]. Any clusters of Camera Angle/Camera Level pairings in a PCA scatter plot have a statistically similar presence in the dataset, providing insights into cinematographic patterns. Lastly, a data visualization of the Camera Angle/Camera Level pairings used sequentially was created for each scene to gain insight into possible temporal patterns.

The next chapter presents the results obtained through the methods described in this chapter. It provides various performance metrics of the model fine-tuning process and obtained models. Then, it presents the outcomes and possible implications of the aforementioned statistical methods, commenting on significance.

Chapter 4

Results

This Chapter presents the obtained results of this research in two phases. Firstly, performance metrics of the performed model fine-tuning are presented, such as training and validation performance as well as model performance on testing data. This is accompanied by an error analysis of the automated classification process. Then, the outcomes of all statistical methods are reported, in the order of descriptive statistics followed by correlational statistics, commenting on implications and significance.

4.1 Model fine-tuning

4.1.1 Initial fine-tuning of the model

The classification report of the testing phase of the best model of the initial fine-tuning is presented in Table 4.1. However, the validation accuracy curves are in fact quite low (around 0.6 - 0.7) (see Appendix A). The misclassification matrices of the best model can be found in Figure 4.1.

Most errors occur close to the diagonal, as expected. However, there are some unexpected errors further away from the diagonal like for example Ground level (true) vs Aerial level (predicted), which are conceptually almost opposite of nature.

4.1.2 Final fine-tuning of the model

The validation accuracy curves over epochs show an improvement with respect to the Dataset BAL training, now being around 0.75 - 0.85. In the misclassification matrices (Figure 4.2) of the best model of this fine-tuning, the errors further away from the diagonals, especially for Camera Level, have almost all been eliminated or diminished.

The classification report of the best model is shown in Table 4.2, while all classification reports and training/validation loss and accuracy curves can be found in Appendix B.

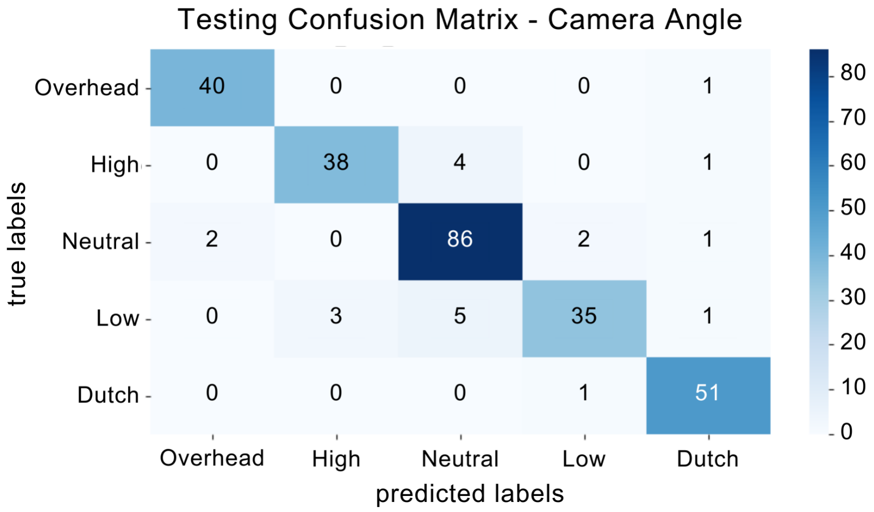
	precision	recall	f1-score	support
Overhead	0.95	0.98	0.96	41
High	0.93	0.88	0.90	43
Neutral	0.91	0.95	0.92	91
Low	0.92	0.80	0.85	44
Dutch	0.93	0.98	0.95	52
accuracy			0.92	271
macro avg	0.93	0.92	0.92	271
weighted avg	0.92	0.92	0.92	271

(a) Camera Angle

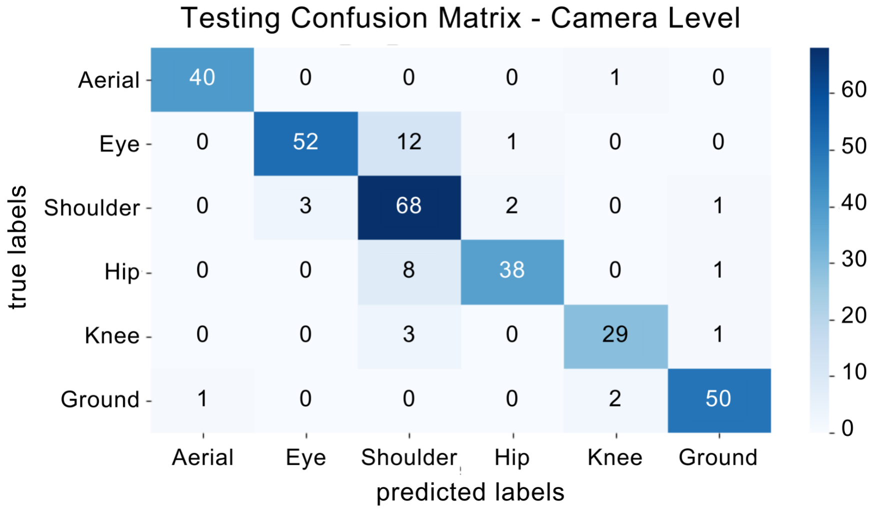
	precision	recall	f1-score	support
Aerial	0.98	0.98	0.98	41
Eye	0.95	0.80	0.87	65
Shoulder	0.75	0.92	0.82	74
Hip	0.93	0.81	0.86	47
Knee	0.91	0.88	0.89	33
Ground	0.94	0.94	0.94	53
accuracy			0.88	313
macro avg	0.91	0.89	0.89	313
weighted avg	0.90	0.88	0.89	313

(b) Camera Level

Table 4.1: Classification reports of the testing phase of the best fine-tuned model on Dataset BAL.



(a) Camera Angle



(b) Camera Level

Figure 4.1: Misclassification matrices of test phase of best model fine-tuned on Dataset BAL.

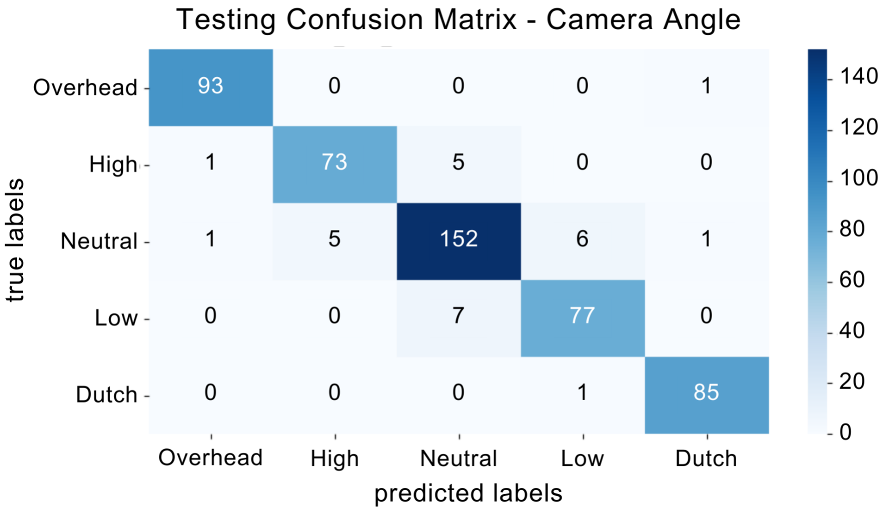
	precision	recall	f1-score	support
Overhead	0.98	0.99	0.98	94
High	0.94	0.93	0.93	84
Neutral	0.93	0.92	0.92	165
Low	0.92	0.92	0.92	84
Dutch	0.98	0.99	0.98	86
accuracy			0.95	513
macro avg	0.95	0.95	0.95	513
weighted avg	0.95	0.95	0.95	513

(a) Camera Angle

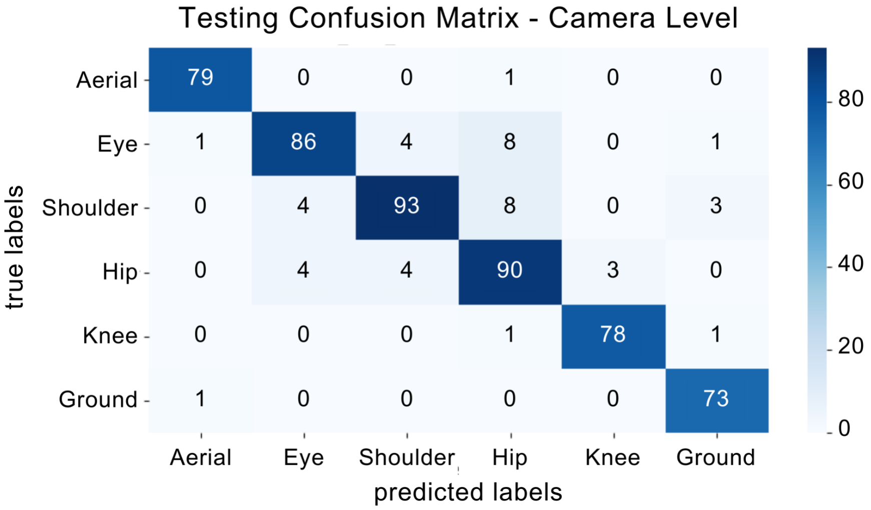
	precision	recall	f1-score	support
Aerial	0.98	0.99	0.98	80
Eye	0.91	0.86	0.89	100
Shoulder	0.92	0.86	0.89	108
Hip	0.83	0.89	0.86	101
Knee	0.96	0.97	0.97	80
Ground	0.94	0.99	0.96	74
accuracy			0.92	543
macro avg	0.92	0.93	0.92	543
weighted avg	0.92	0.92	0.92	543

(b) Camera Level

Table 4.2: Classification reports of the testing phase of the best fine-tuned model on Dataset AUG.



(a) Camera Angle



(b) Camera Level

Figure 4.2: Misclassification matrices of test phase of best model fine-tuned on Dataset AUG.

The obtained precision, recall and accuracy scores are between 83-98 %, of which most above 90 percent. Considering the complexity and heterogeneity of a data population such as film frames and the limited quantity and quality (due to ambiguity) of training data, these results are deemed acceptable.

4.1.3 Error analysis

When examining the misclassified images of the testing phase of the final fine-tuned model, some primary causes were identified. Some hip level shots were classified as eye level and vice versa, typically because there was a non-standing subject in the frame, or an unclear geometry. Other incorrectly labeled frames were often misclassified due to lack of context, cluttered scenes or motion blur.

4.2 Statistical analysis

4.2.1 Descriptive statistics

The distributions of Camera Angle and level both show significant differences amongst classes. Neutral angles are strongly predominant, accounting for 50.80% of all frames, while the secondmost occurring Camera Angle Low accounts for 21.91%. Overhead angle appears the least, in only 2.29% of frames. In the Camera Level distribution, however, the top two occurring classes are much closer together. The most frequent being Shoulder (38.63%), followed by Eye (33.61%). Ground is the least represented at 2.85% occurrence. The occurrence counts for both features are presented in Table 4.3.

Camera Angle	Count	%
Neutral	11407	50.80
Low	4919	21.91
High	4249	18.92
Dutch	1365	6.08
Overhead	514	2.29

(a) Camera Angle

Camera Level	Count	%
Shoulder	8673	38.63
Eye	7547	33.61
Hip	4094	18.23
Aerial	837	3.73
Knee	664	2.96
Ground	639	2.85

(b) Camera Level

Table 4.3: Overview of the occurrence counts of Camera Angle and Camera Level.

The average scene duration in the dataset is 3.74 minutes, with a median duration of 3.28 minutes. The distribution of scene duration is heavily right-skewed, as can be seen in Figure 4.3. The shortest scene is 58 seconds long, while the longest runs for 9.15 minutes. In fact, the longest scene falls outside of the upper bound of median +- 1.5 * IQR (inter-quartile range), but the right-skewed nature of the dataset suggests that it might be part of the distribution rather than a true anomaly. The full list of descriptive statistics for scene duration, including all quartile partitions, can be found in Table 4.4.

	Seconds	Minutes
mean	224.54	3.74
std	91.77	1.53
min	58.00	0.97
25%	155.25	2.59
median	197.00	3.28
75%	278.25	4.64
max	549.00	3.28

Table 4.4: Descriptive statistics of scene duration.

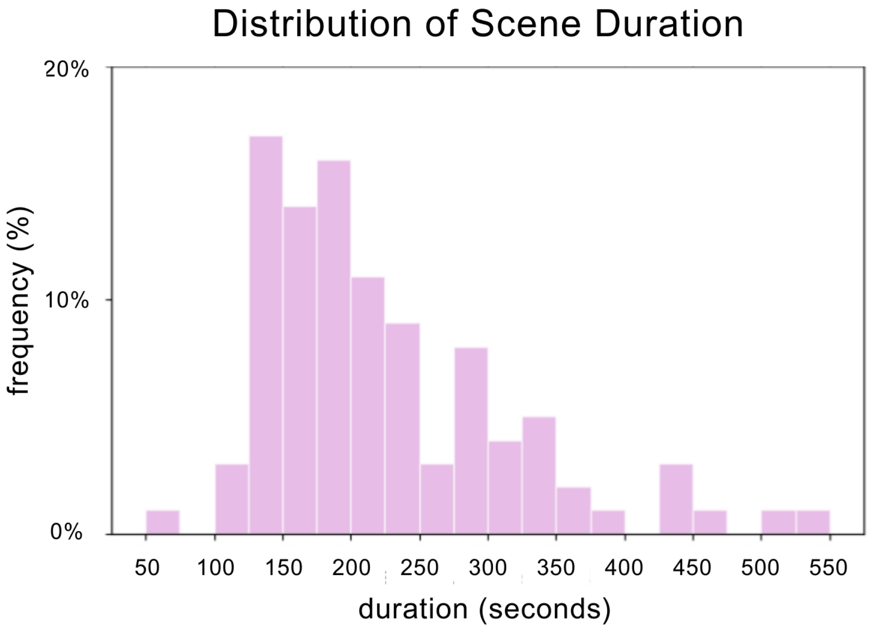


Figure 4.3: The frequency distribution of scene duration.

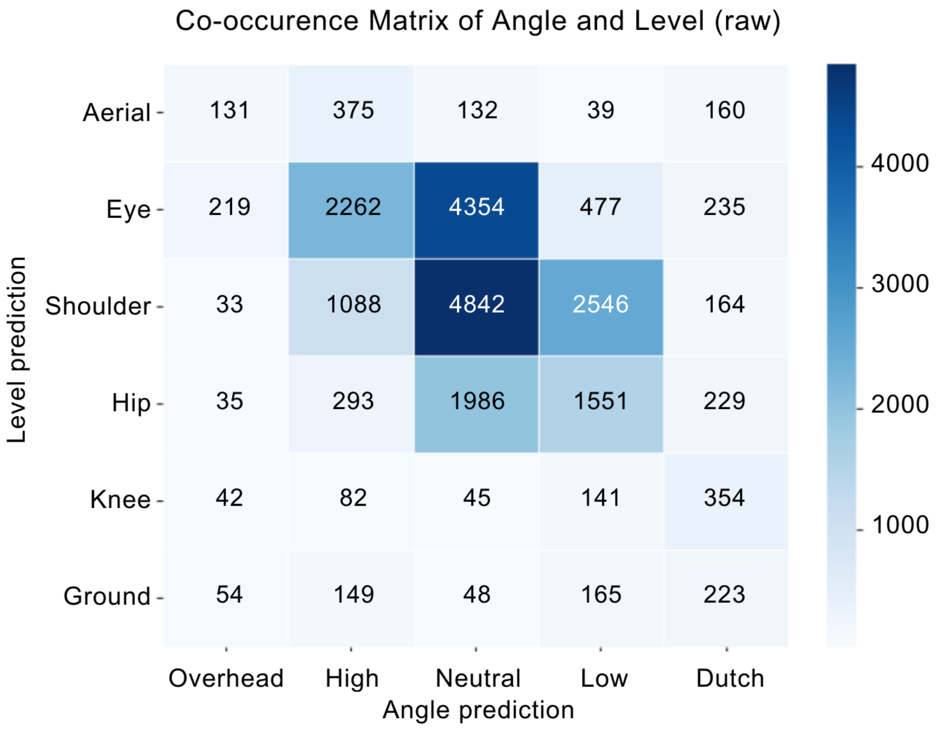
4.2.2 Correlational statistics

Two co-occurence matrices were computed for Camera Angle/Camera Level pairs, one with raw counts and one normalized for each level class, as shown in Figure 4.4. For the raw counts, it is clear that neutral angle shots at shoulder or eye level are the two very dominant presences in the dataset, accounting for for 40.95 percent of frames together. To account for imbalances in overall frequency among Camera Levels, the co-occurrence matrix was computed normalized per level class. This reveals that the extremer classes of either feature (Aerial, Knee and ground levels - overhead, high, low and Dutch) are often paired amongst each other. The more neutral level classes (eye, shoulder, hip) are very rarely paired with extreme angles such as overhead or Dutch.

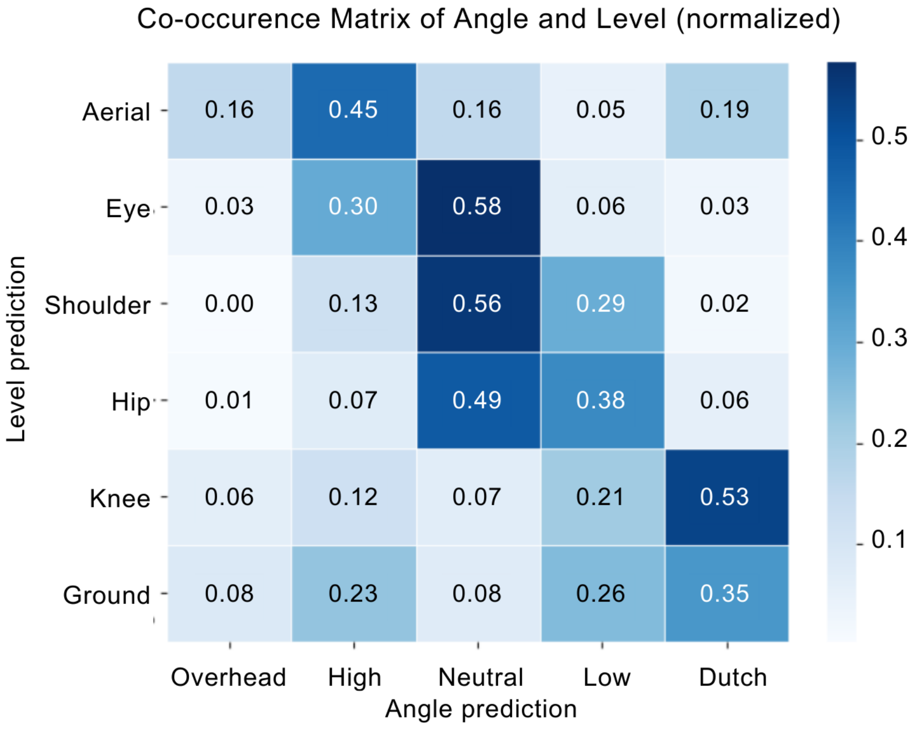
It is unsurprising, then, that the Chi-Squared test of independence produced a highly significant p-value ($p \approx 0.00$) indicating dependence between Camera Angle and Camera Level. Cramér's V yielded a value of 0.31, suggesting a moderate association between the two features. A heatmap of the standardized residuals, shown in Figure 4.5, further reinforces the idea that angle-level pairings follow established visual conventions. The upper left part of the heatmap (the cells corresponding to aerial until hip levels vs overhead until low angles) contains positive residuals in a diagonal fashion, likely due to the subject of a shot often being (the face(s) of) one or more standing or sitting human subjects. Higher heights with a high angle, eye level with an (almost) neutral level and lower heights with a low angle all potentially frame a subject's face.

PCA applied to one-hot coded Camera Angles/Camera Level pairs (using slight jitter for improved readability of overlapping datapoints), though quite dispersed overall, shows some clustered datapoints, suggesting some groups of angle/level pairs appear in the dataset in a statistically similar way (Figure 4.6). The inter-level spacings of a single angle have a similar pattern across all angles. Their aerial, knee and ground level are clustered together not far from their hip level, while their eye and shoulder levels are spaced further away in opposite directions along a diagonal. Furthermore, the levels corresponding to an overhead or Dutch angle are clustered by level. It should be noted that the PCA was applied to the dataset containing data on frames of all scenes concatenated, and thus has no context of scene or scene duration, identifying latent structures purely based on each angle/level pair's statistical properties within the dataset.

Lastly, Camera Angle, Camera Level and Camera Angle/Camera Level pairs plotted over time for each scene show a vast variety in the temporal domain. Some scenes contained almost exclusively neutral angles and/or shoulder/eye level shots. Others contain a wide variety of pairings. Multiple consecutive frames of the same Angle/Level pair could belong to the same shot, but not necessarily. Conversely, consecutive Angle/Level pairs changing rapidly are unlikely (but not impossible) to belong to the same shot. The scenes containing a wide variety of pairings often also had pairings changing very rapidly over time, consistently throughout most or all of the scene. This could be a stylistic pattern, though it possibly indicates noisy data instead, perhaps due to those particular scenes being underrepresented in the training data. Some examples of the Camera Angle/Camera Level visualization are shown in Figure 4.7, the full visualizations can be found in Appendix C.



(a) Raw counts



(b) Normalized per level

Figure 4.4: Co-occurence matrices of Camera Angle/Camera Level, raw counts and normalized per level.

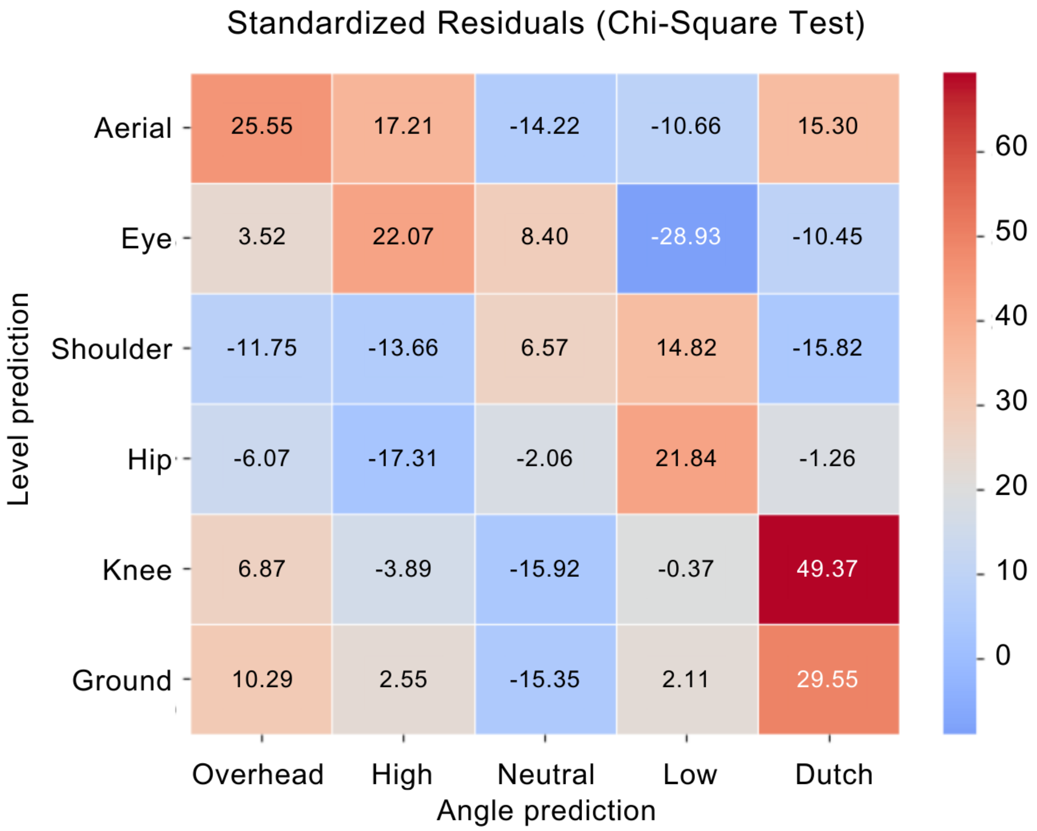


Figure 4.5: A heatmap of standardized residuals of the Chi-Squared test.

The conclusions purported by these results will be emphasized and commented on in the next chapter, after a summary of the steps taken to obtain these results. Then, suggestions and ideas for future research brought forth by this research are outlined.

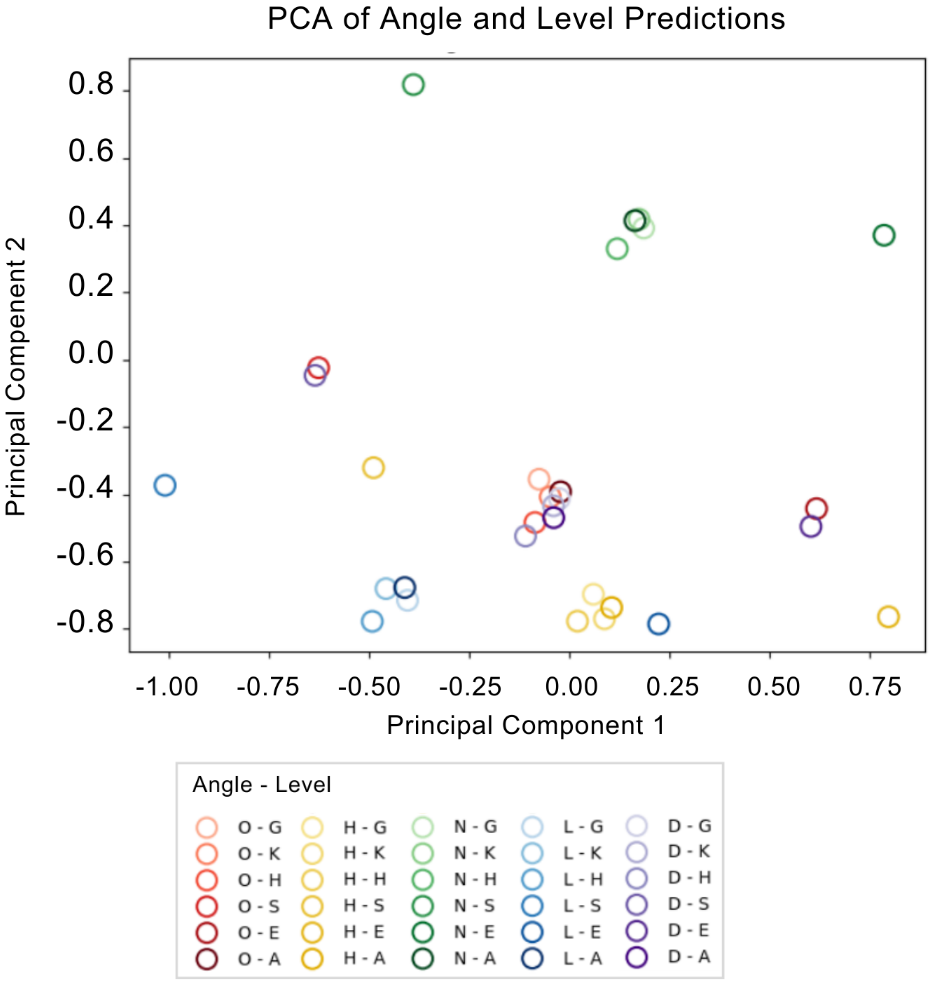


Figure 4.6: PCA on one-hot encoded Camera Angle and Camera Level.

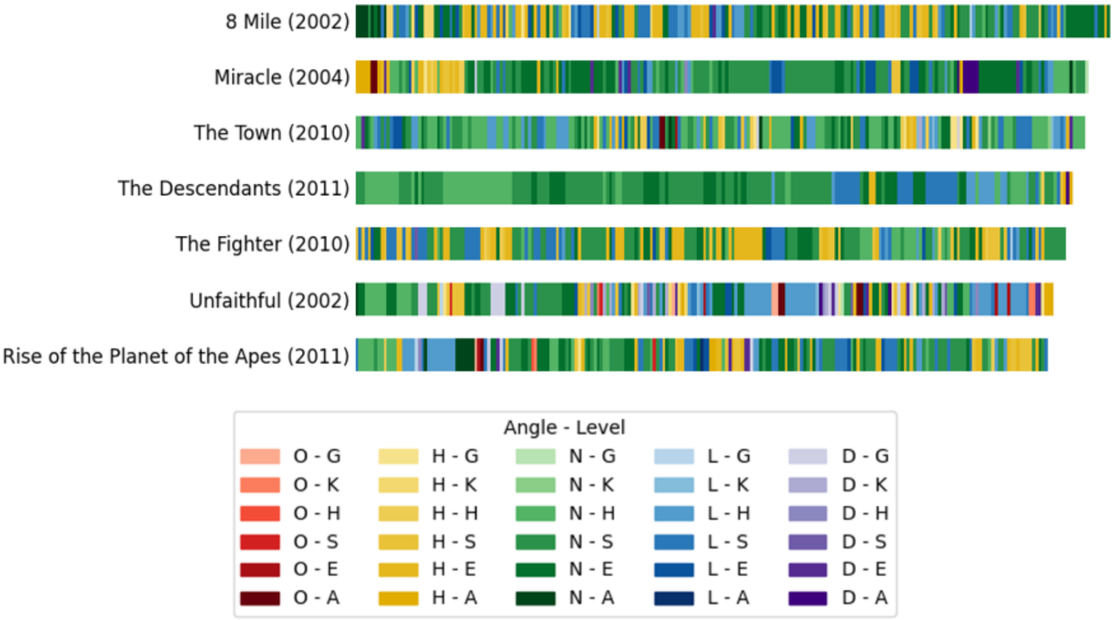


Figure 4.7: A subset of temporal visualizations of scenes by Camera Angle/Camera Level pairing.

Conclusions and future work

The goal of this research was to correlate the use of Camera Angle and Camera Level as stylistic features of scenes in contemporary and popular empathy-eliciting Hollywood films. Empirical data on Camera Angle and Camera Level was automatically extracted from 22,454 film frames with the use of a CNN model. Through a custom data augmentation pipeline, training data was able to fine-tune a pre-existing CNN model with satisfactory validation performance, precision, recall and accuracy.

The most common Camera Angle and Camera Level classes identify stylistic tendencies. It is revealed that half (50.80%) of all film frames are shot at a Neutral Angle, a uniquely strong dominance over other Camera Angle classes. Amongst Camera Level classes, Shoulder Level (38.63%) and Eye Level (33.61%) dominate together. Overhead Angle (2.29%) and Ground Level (2.85%) are rare. When considering Camera Angle / Camera Level pairings it is shown that these dominant classes often appear together. Overall, the statistical analysis indicates a moderate but highly significant correlation between Camera Angle and Camera Level, with a p-value of almost 0 for the Chi-Squared test and Cramér's V being 0.31. There are clear statistical patterns in the use of level-angle pairings. More neutral Level classes (Eye, Shoulder and Hip Level) are predominantly paired with High, Neutral or Low angles. Similarly, extremer classes of either feature (Aerial, Knee and Ground Level, Overhead and Dutch Angle) tend to be paired amongst each other. Other pairing conventions include those that naturally align with framing human subjects.

Limitations

The fine-tuning of the model performed to ensure acceptable performance on the full sample of film frames was done with a relatively small and not perfectly balanced training dataset, possibly introducing erroneous biases. Validation Accuracy of 75-85% was deemed acceptable in the scope of this research but could benefit from higher quantity and quality training data. The quality was possibly limited due to ambiguity. The frequent nature of Eye Level shots suggests the need for an additional transitionary level class in between Eye Level and Aerial Level.

Future research

The Convolution Neural Network specially trained to predict the Camera Angle and Camera Level used in single film frames and the annotated film frame dataset brought forth by this research may serve as resources for future research in this field. In further research employing the automated extraction of Camera Angle and Camera Level, model performance could be improved by fine-tuning with higher quantity and quality Ground Truth data. The classes of both Camera Angle and Camera Level could be expanded with some additional transitionary classes to reduce ambiguity in the data. Furthermore, a statistical analysis of the use of Camera Angle and Camera Level temporally on a large amount of scenes could provide insights into more complex patterns and interdependencies in the use of the two features.

Future correlational experiments could be conducted on film-specific empirical data of viewers' emotional responses and formal camera features. As the body of research in this field grows, research could be done to investigate the accuracy with which emotional responses can be predicted based on analysis of content and formal features combined. Possible applications of such predictions could be explored to benefit various aspects of filmmaking, audience suitability rating, video advertising, content suggestion algorithms, content generation and more.

Appendix A

Initial fine-tuning

In this Appendix, the performances of the initial fine-tuning rounds are presented for four different hyperparameter configurations. The following sections provide plots of their training losses, training accuracies, validation accuracies, classification reports, and testing phase prediction/ground truth confusion matrices.

Training loss

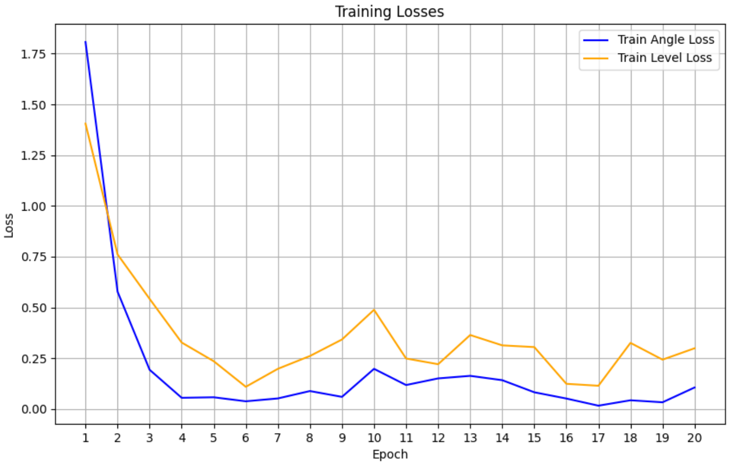


Figure A.1.1 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

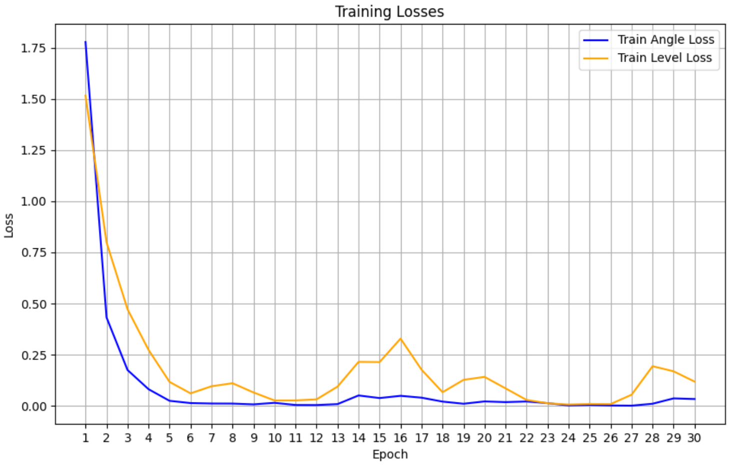


Figure A.1.2 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.

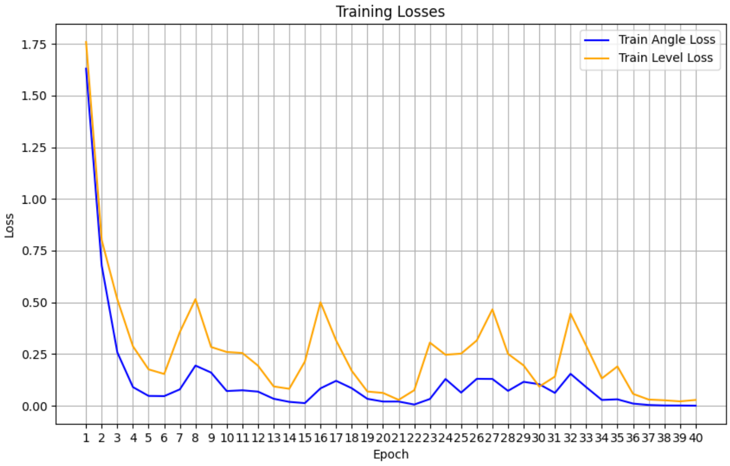


Figure A.1.3 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

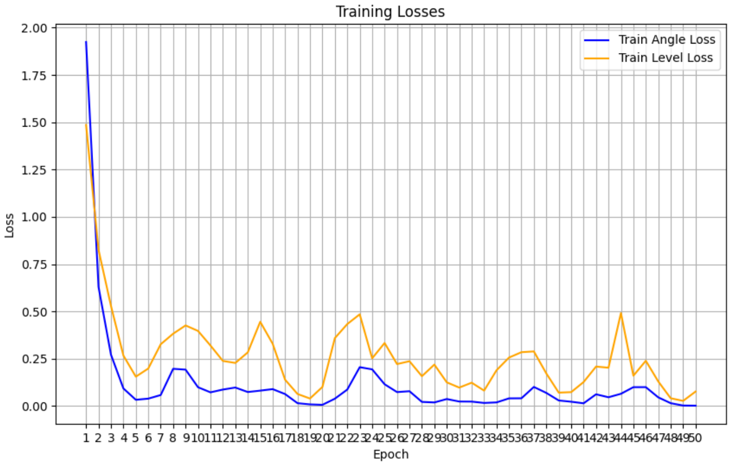


Figure A.1.4 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Training Accuracy and Validation Accuracy

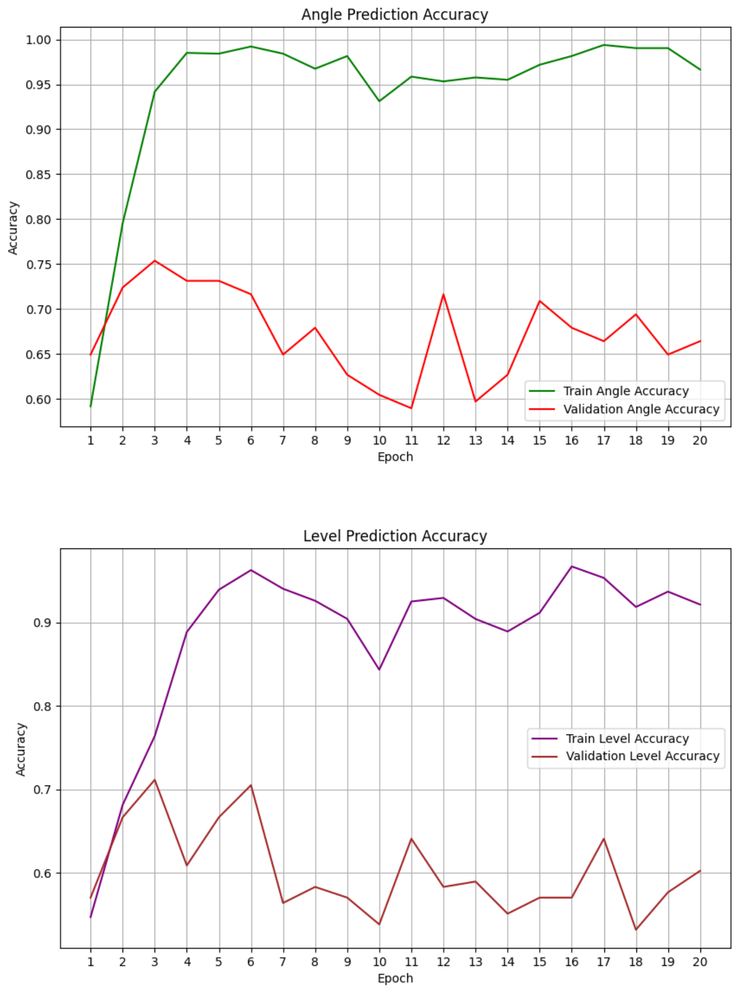


Figure A.2.1 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

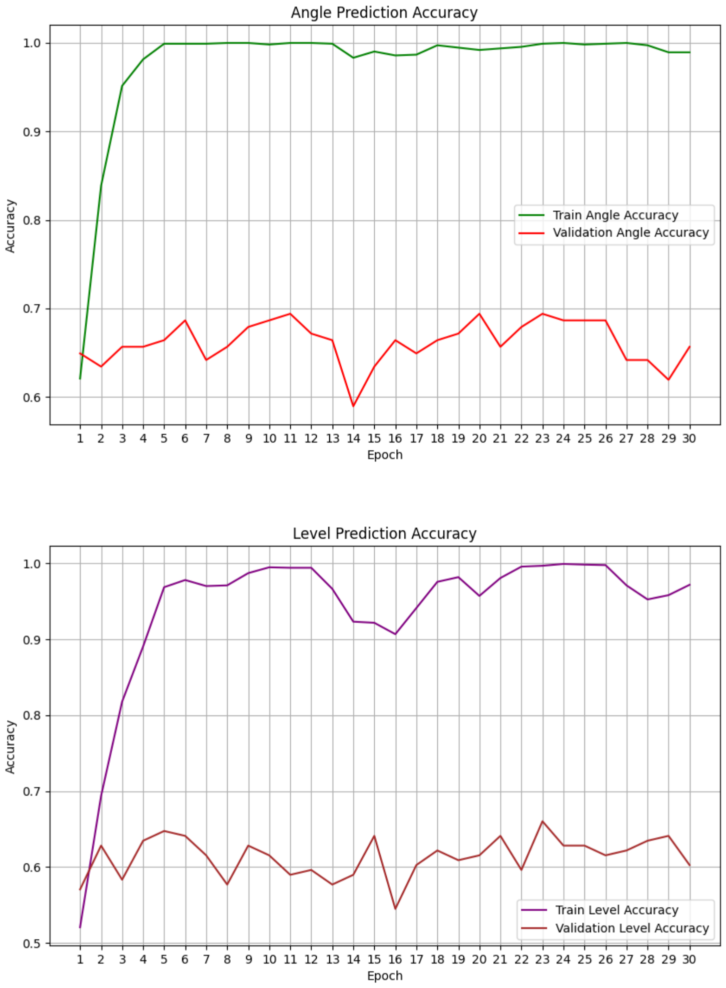


Figure A.2.2 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=5e-4.

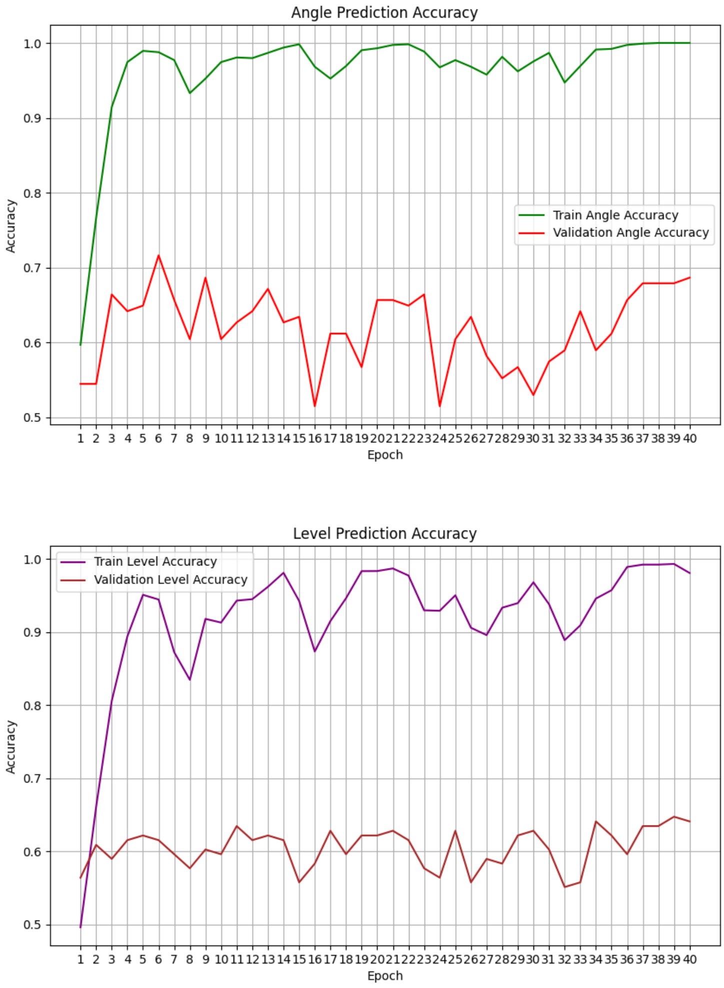


Figure A.2.3 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

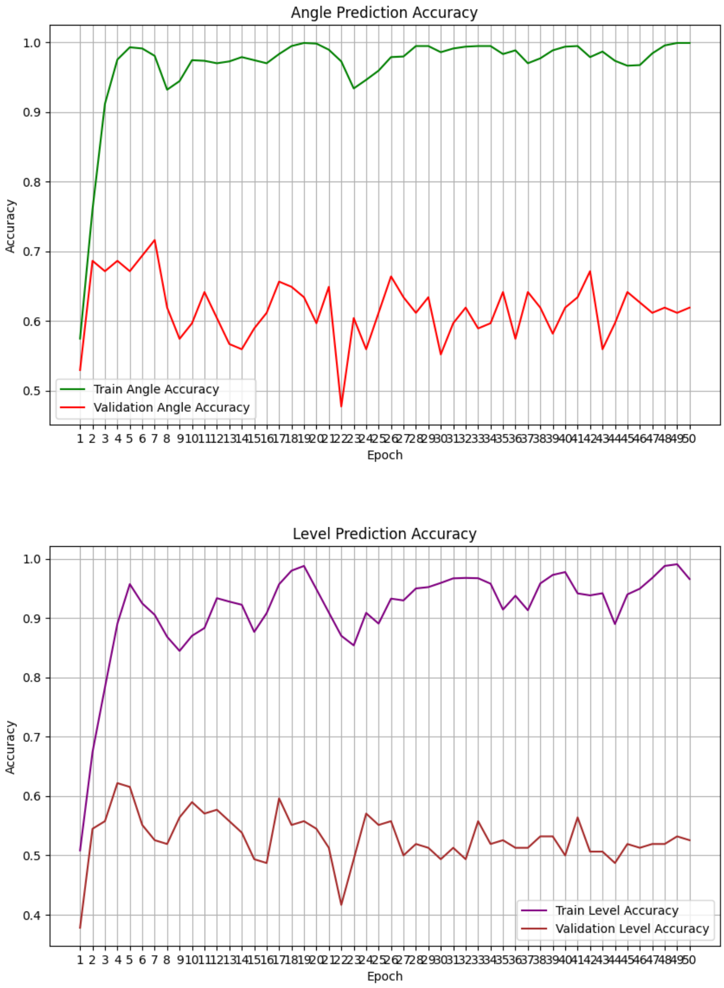
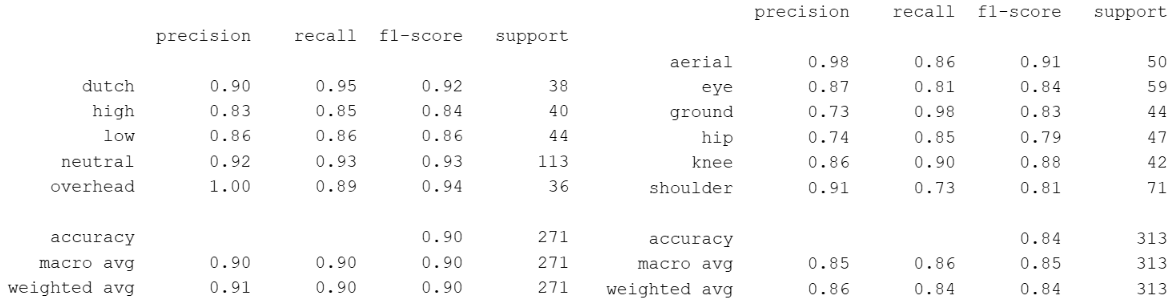


Figure A.2.4 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

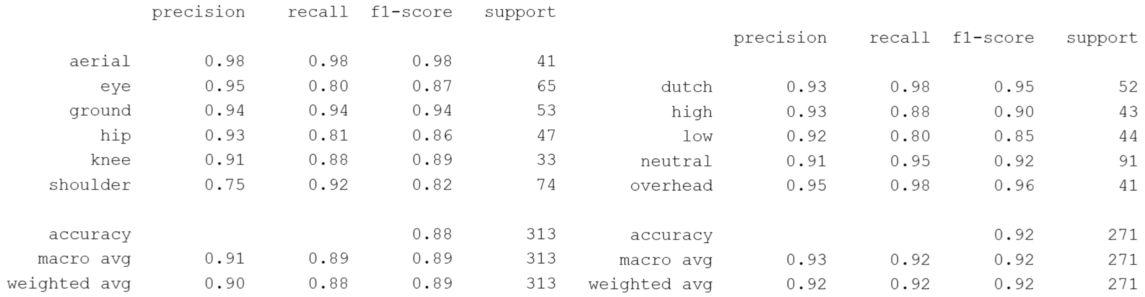
Testing Phase Classification Report



```
              precision    recall  f1-score   support

       dutch       0.90      0.95      0.92        38
        high       0.83      0.85      0.84        40
         low       0.86      0.86      0.86        44
     neutral       0.92      0.93      0.93       113
    overhead       1.00      0.89      0.94        36

    accuracy                           0.90       271
   macro avg       0.90      0.90      0.90       271
weighted avg       0.91      0.90      0.90       271
```

```
              precision    recall  f1-score   support

      aerial       0.98      0.86      0.91        50
         eye       0.87      0.81      0.84        59
      ground       0.73      0.98      0.83        44
         hip       0.74      0.85      0.79        47
        knee       0.86      0.90      0.88        42
    shoulder       0.91      0.73      0.81        71

    accuracy                           0.84       313
   macro avg       0.85      0.86      0.85       313
weighted avg       0.86      0.84      0.84       313
```

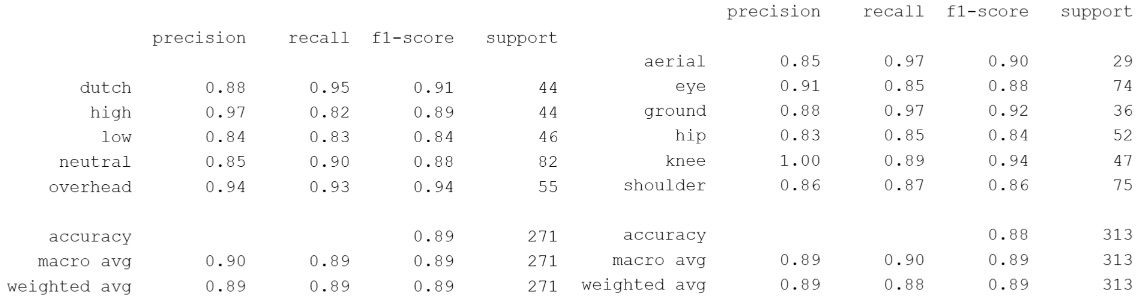
Figure A.3.1 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.



```
              precision    recall  f1-score   support

      aerial       0.98      0.98      0.98        41
         eye       0.95      0.80      0.87        65
      ground       0.94      0.94      0.94        53
         hip       0.93      0.81      0.86        47
        knee       0.91      0.88      0.89        33
    shoulder       0.75      0.92      0.82        74

    accuracy                           0.88       313
   macro avg       0.91      0.89      0.89       313
weighted avg       0.90      0.88      0.89       313
```

```
              precision    recall  f1-score   support

       dutch       0.93      0.98      0.95        52
        high       0.93      0.88      0.90        43
         low       0.92      0.80      0.85        44
     neutral       0.91      0.95      0.92        91
    overhead       0.95      0.98      0.96        41

    accuracy                           0.92       271
   macro avg       0.93      0.92      0.92       271
weighted avg       0.92      0.92      0.92       271
```

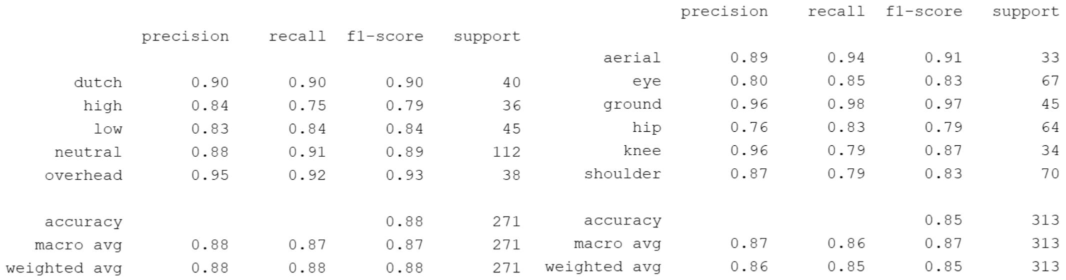
Figure A.3.2 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.



```
              precision    recall  f1-score   support

       dutch       0.88      0.95      0.91        44
        high       0.97      0.82      0.89        44
         low       0.84      0.83      0.84        46
     neutral       0.85      0.90      0.88        82
    overhead       0.94      0.93      0.94        55

    accuracy                           0.89       271
   macro avg       0.90      0.89      0.89       271
weighted avg       0.89      0.89      0.89       271
```

```
              precision    recall  f1-score   support

      aerial       0.85      0.97      0.90        29
         eye       0.91      0.85      0.88        74
      ground       0.88      0.97      0.92        36
         hip       0.83      0.85      0.84        52
        knee       1.00      0.89      0.94        47
    shoulder       0.86      0.87      0.86        75

    accuracy                           0.88       313
   macro avg       0.89      0.90      0.89       313
weighted avg       0.89      0.88      0.89       313
```

Figure A.3.3 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.



```
              precision    recall  f1-score   support

       dutch       0.90      0.90      0.90        40
        high       0.84      0.75      0.79        36
         low       0.83      0.84      0.84        45
     neutral       0.88      0.91      0.89       112
    overhead       0.95      0.92      0.93        38

    accuracy                           0.88       271
   macro avg       0.88      0.87      0.87       271
weighted avg       0.88      0.88      0.88       271
```

```
              precision    recall  f1-score   support

      aerial       0.89      0.94      0.91        33
         eye       0.80      0.85      0.83        67
      ground       0.96      0.98      0.97        45
         hip       0.76      0.83      0.79        64
        knee       0.96      0.79      0.87        34
    shoulder       0.87      0.79      0.83        70

    accuracy                           0.85       313
   macro avg       0.87      0.86      0.87       313
weighted avg       0.86      0.85      0.85       313
```

Figure A.3.4 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Testing Phase Confusion Matrices

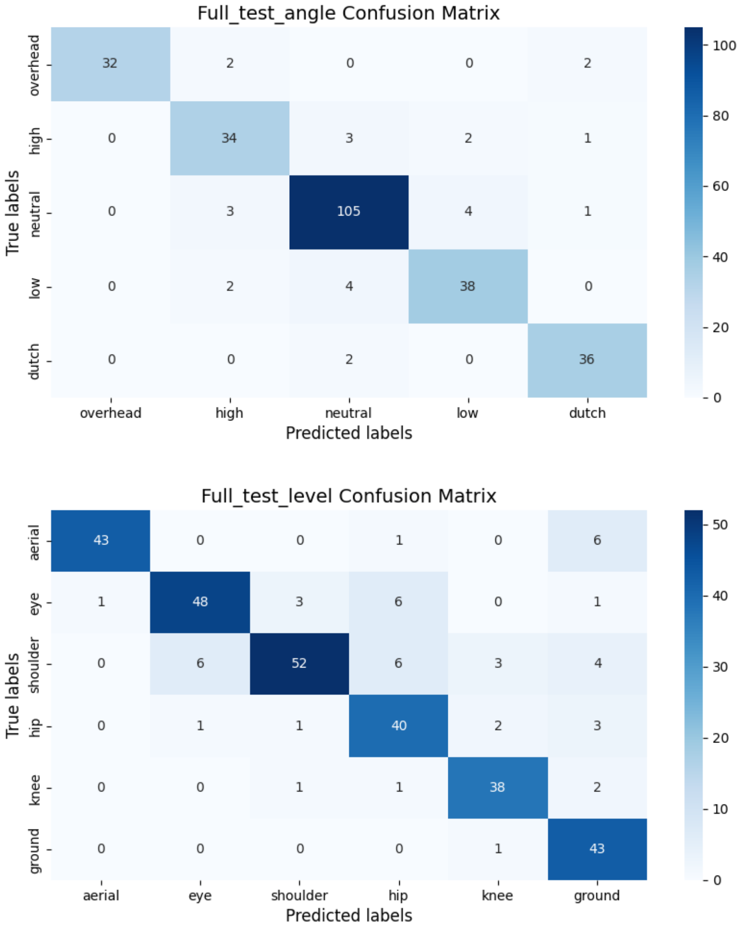


Figure A.4.1 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

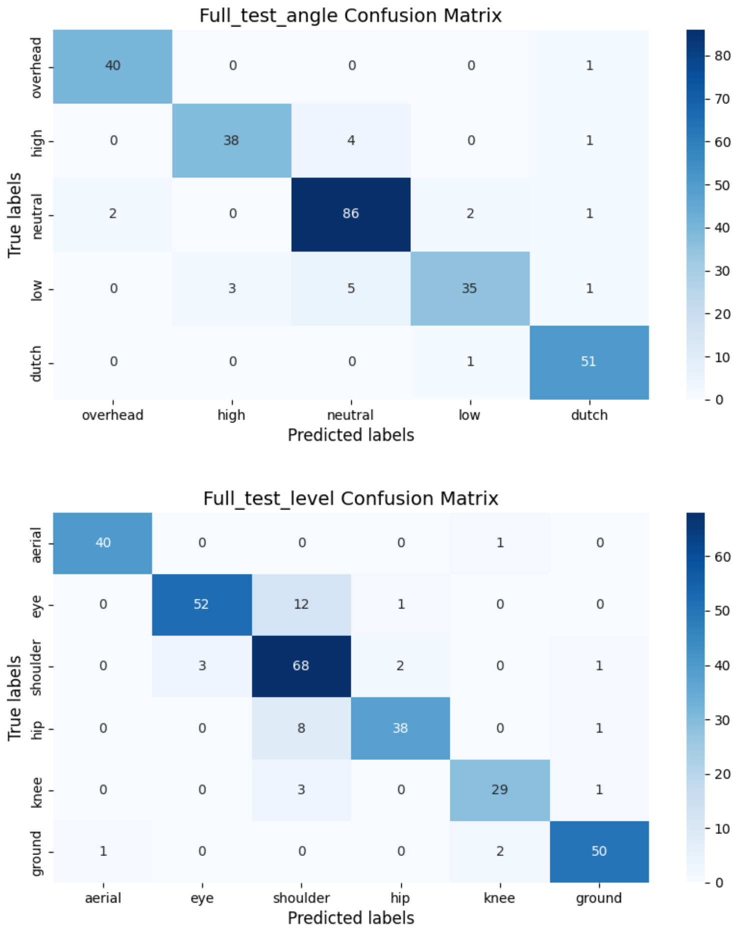
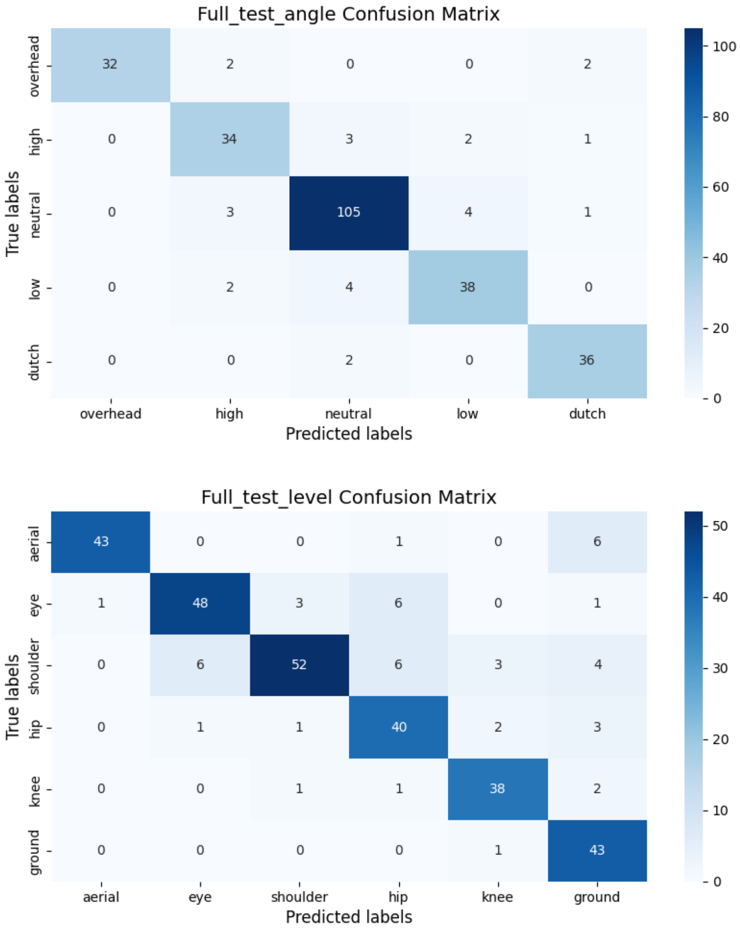


Figure A.4.2 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.



	overhead	high	neutral	low	dutch
overhead	32	2	0	0	2
high	0	34	3	2	1
neutral	0	3	105	4	1
low	0	2	4	38	0
dutch	0	0	2	0	36

	aerial	eye	shoulder	hip	knee	ground
aerial	43	0	0	1	0	6
eye	1	48	3	6	0	1
shoulder	0	6	52	6	3	4
hip	0	1	1	40	2	3
knee	0	0	1	1	38	2
ground	0	0	0	0	1	43

Figure A.4.3 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

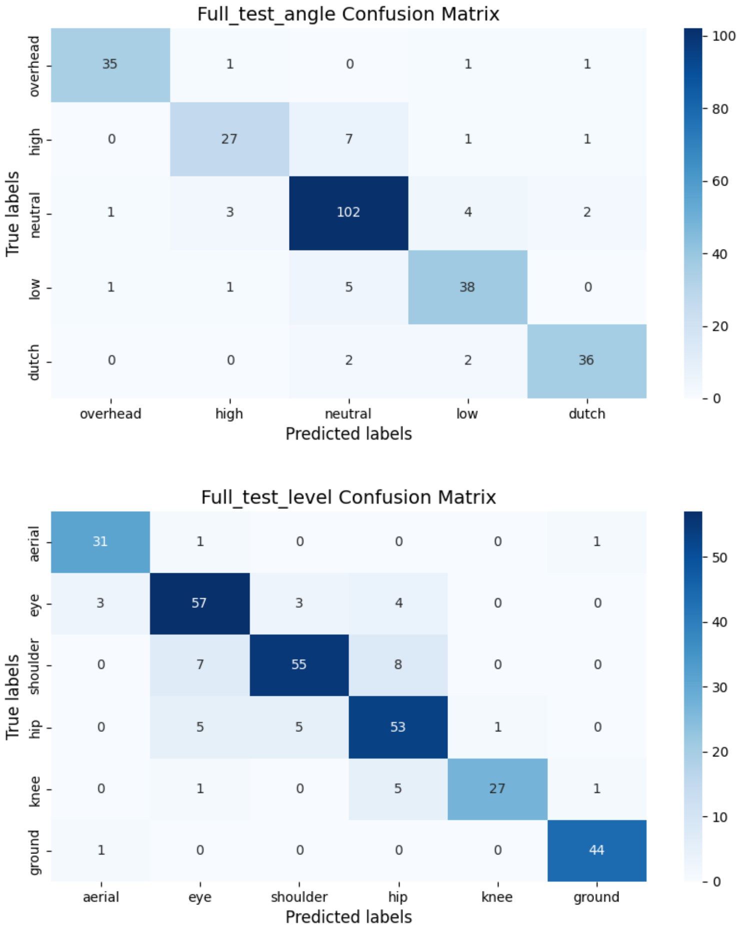


Figure A.4.4 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Appendix B

Final Fine-tuning

In this Appendix, the performances of the final fine-tuning rounds are presented for four different hyperparameter configurations. The following sections provide plots of their training losses, training accuracies, validation accuracies, classification reports, and testing phase prediction/ground truth confusion matrices.

Training loss

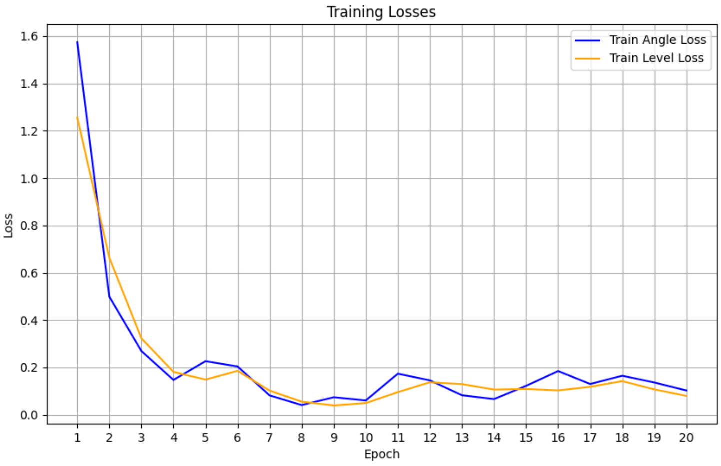


Figure B.1.1 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

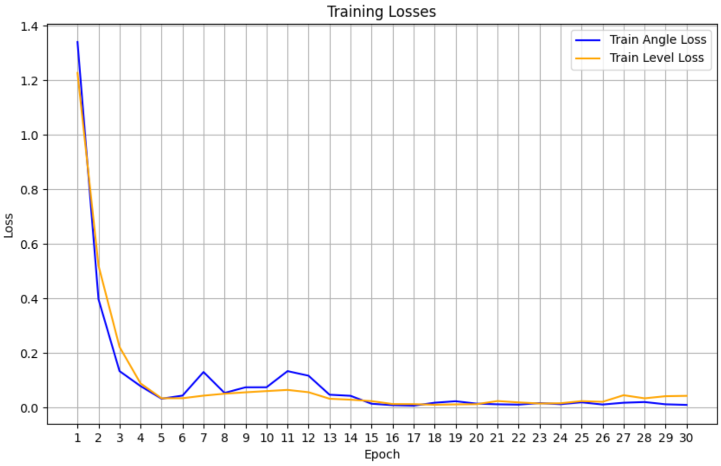


Figure B.1.2 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.

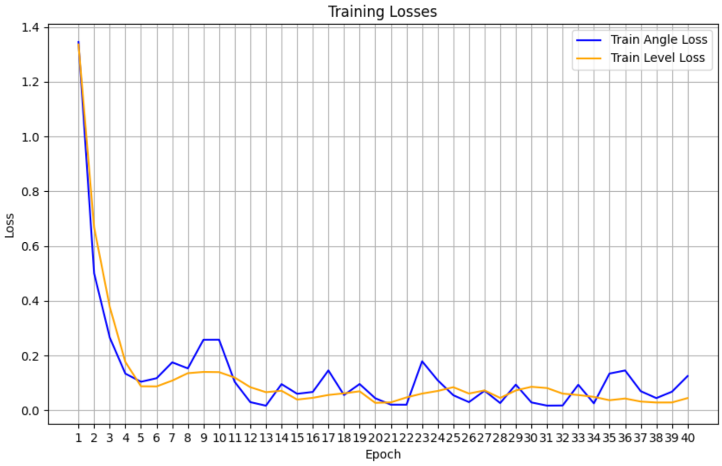


Figure B.1.3 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

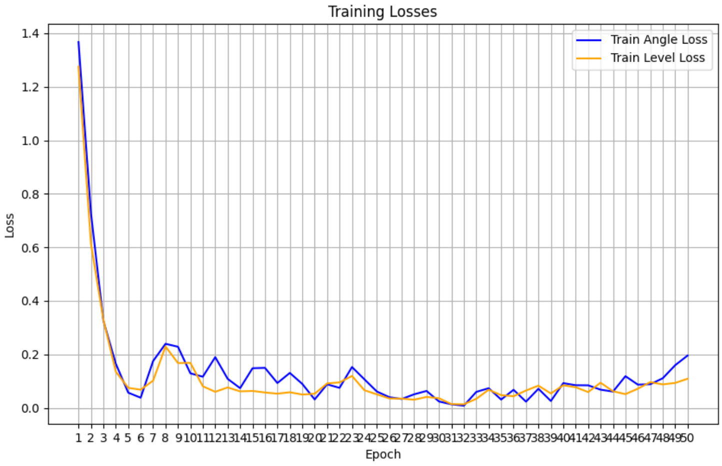


Figure B.1.4 Training loss curve over epochs for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Training Accuracy and Validation Accuracy

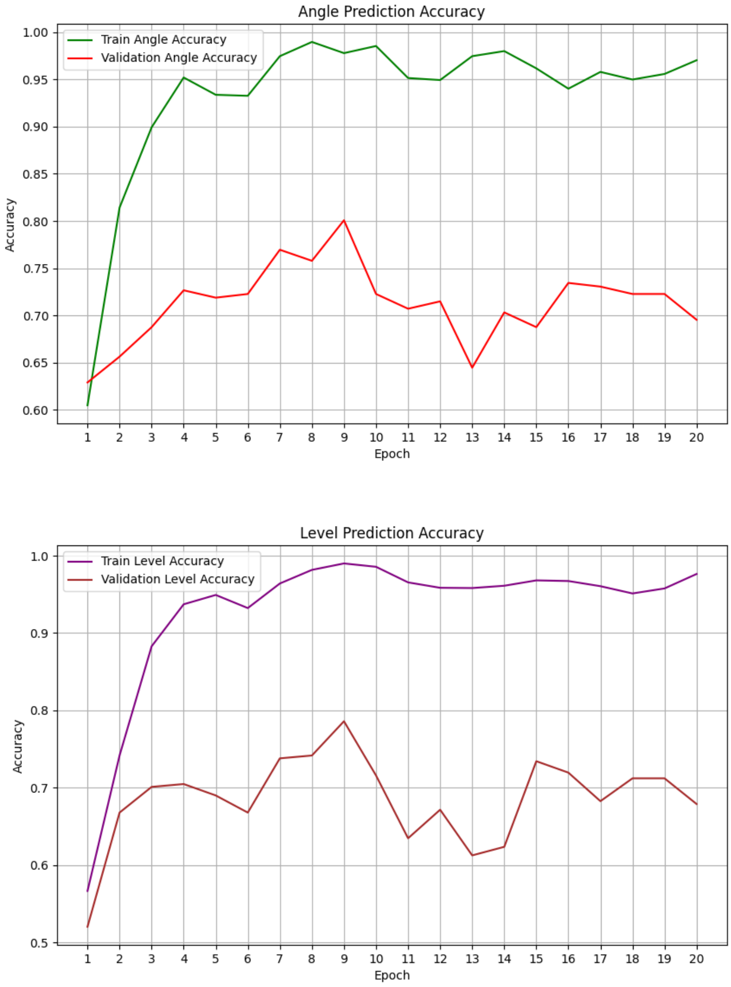


Figure B.2.1 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

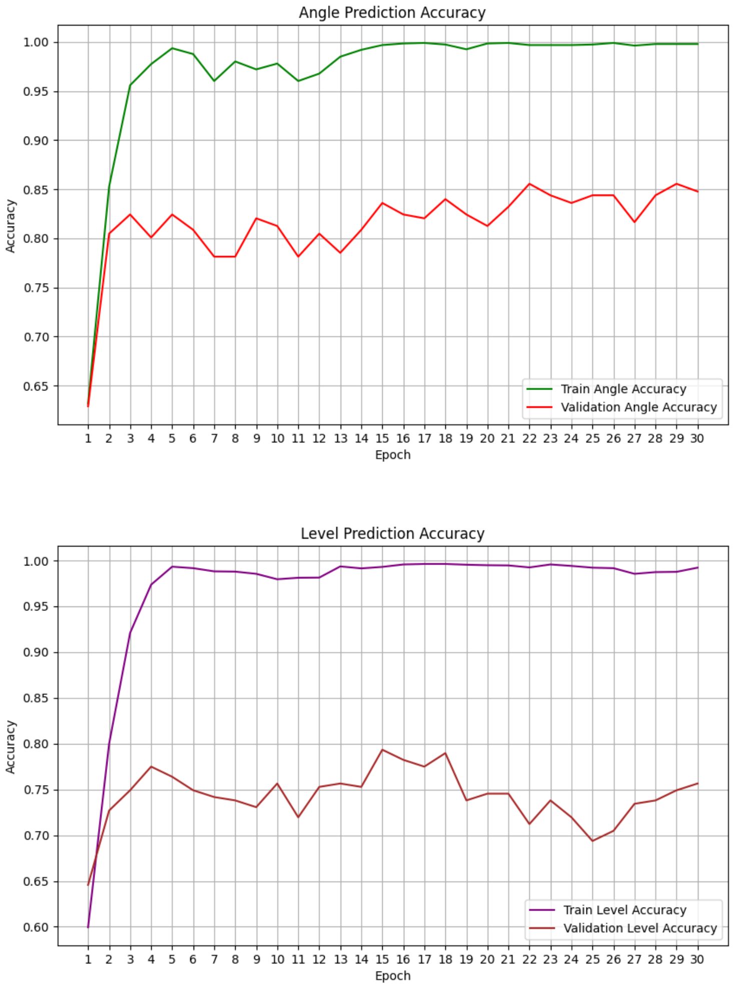


Figure B.2.2 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=5e-4.

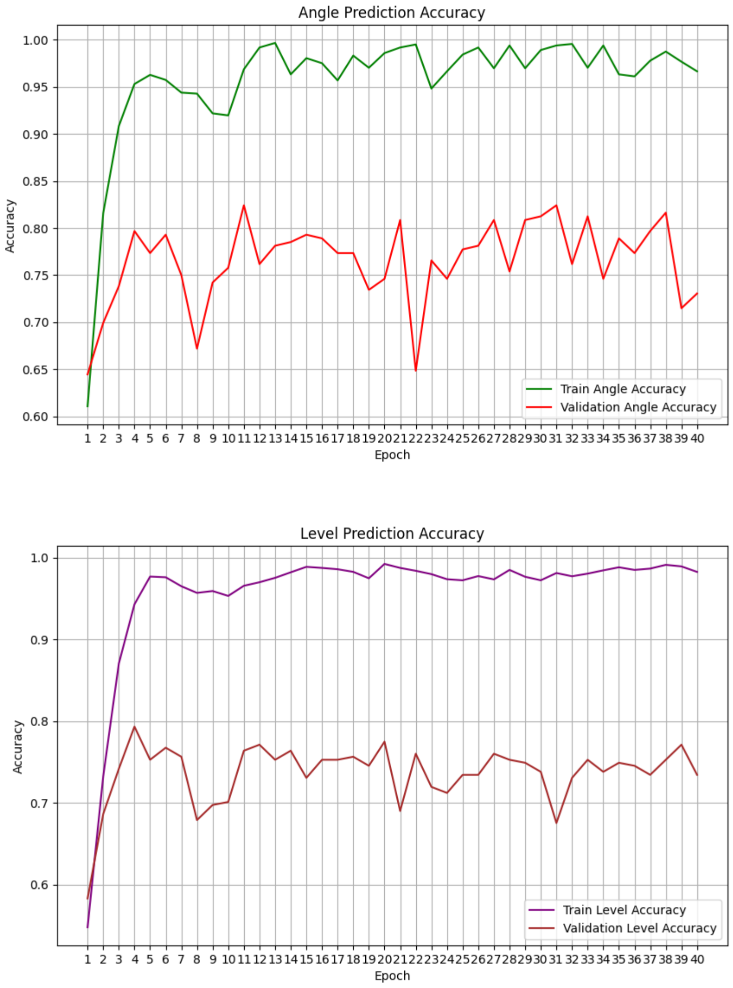


Figure B.2.3 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

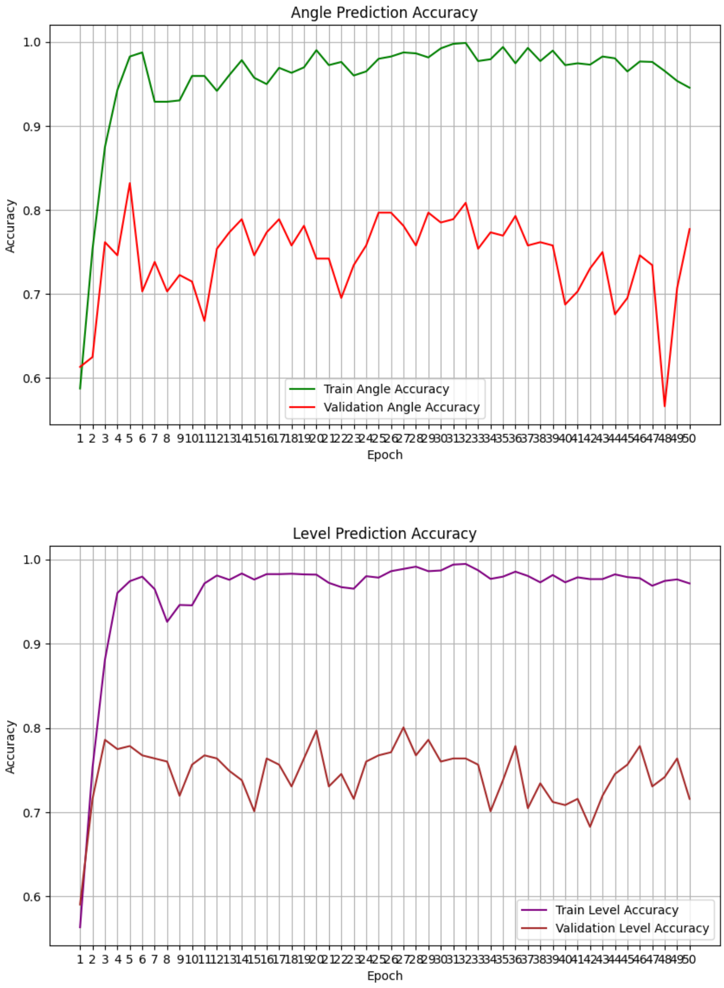
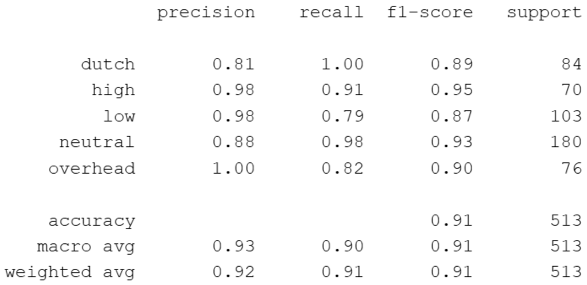
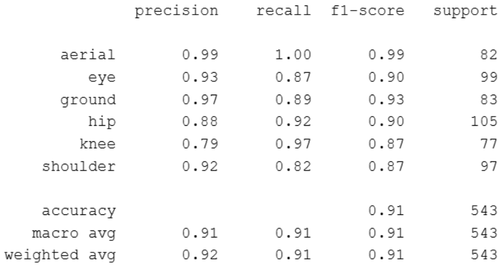


Figure B.2.4 Training Accuracy and Validation Accuracy curve over epochs for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Testing Phase Classification Report

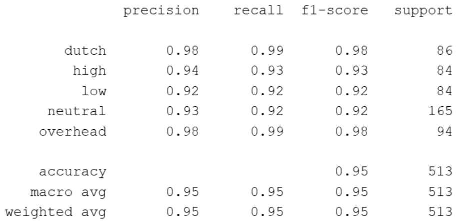


```
              precision    recall  f1-score   support

       dutch       0.81      1.00      0.89        84
        high       0.98      0.91      0.95        70
         low       0.98      0.79      0.87       103
     neutral       0.88      0.98      0.93       180
    overhead       1.00      0.82      0.90        76

    accuracy                           0.91       513
   macro avg       0.93      0.90      0.91       513
weighted avg       0.92      0.91      0.91       513
```

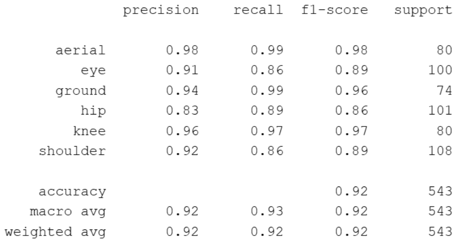


```
              precision    recall  f1-score   support

      aerial       0.99      1.00      0.99        82
         eye       0.93      0.87      0.90        99
      ground       0.97      0.89      0.93        83
         hip       0.88      0.92      0.90       105
        knee       0.79      0.97      0.87        77
    shoulder       0.92      0.82      0.87        97

    accuracy                           0.91       543
   macro avg       0.91      0.91      0.91       543
weighted avg       0.92      0.91      0.91       543
```

Figure B.3.1 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

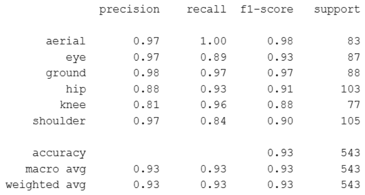


```
              precision    recall  f1-score   support

       dutch       0.98      0.99      0.98        86
        high       0.94      0.93      0.93        84
         low       0.92      0.92      0.92        84
     neutral       0.93      0.92      0.92       165
    overhead       0.98      0.99      0.98        94

    accuracy                           0.95       513
   macro avg       0.95      0.95      0.95       513
weighted avg       0.95      0.95      0.95       513
```

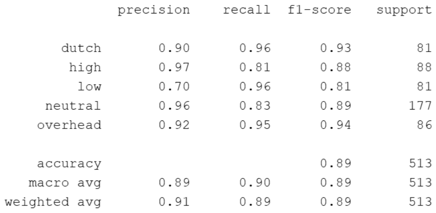


```
              precision    recall  f1-score   support

      aerial       0.98      0.99      0.98        80
         eye       0.91      0.86      0.89       100
      ground       0.94      0.99      0.96        74
         hip       0.83      0.89      0.86       101
        knee       0.96      0.97      0.97        80
    shoulder       0.92      0.86      0.89       108

    accuracy                           0.92       543
   macro avg       0.92      0.93      0.92       543
weighted avg       0.92      0.92      0.92       543
```

Figure B.3.2 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.

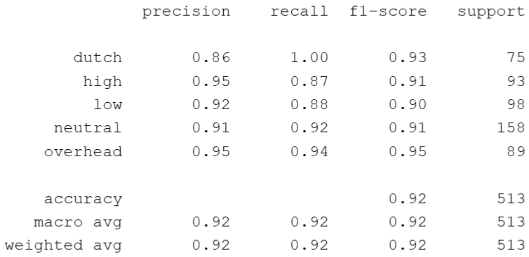


```
              precision    recall  f1-score   support

      aerial       0.97      1.00      0.98        83
         eye       0.97      0.89      0.93        87
      ground       0.98      0.97      0.97        88
         hip       0.88      0.93      0.91       103
        knee       0.81      0.96      0.88        77
    shoulder       0.97      0.84      0.90       105

    accuracy                           0.93       543
   macro avg       0.93      0.93      0.93       543
weighted avg       0.93      0.93      0.93       543
```

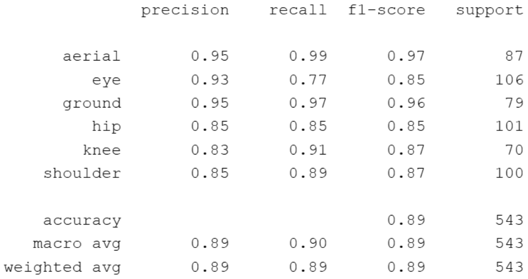


```
              precision    recall  f1-score   support

       dutch       0.90      0.96      0.93        81
        high       0.97      0.81      0.88        88
         low       0.70      0.96      0.81        81
     neutral       0.96      0.83      0.89       177
    overhead       0.92      0.95      0.94        86

    accuracy                           0.89       513
   macro avg       0.89      0.90      0.89       513
weighted avg       0.91      0.89      0.89       513
```

Figure B.3.3 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.



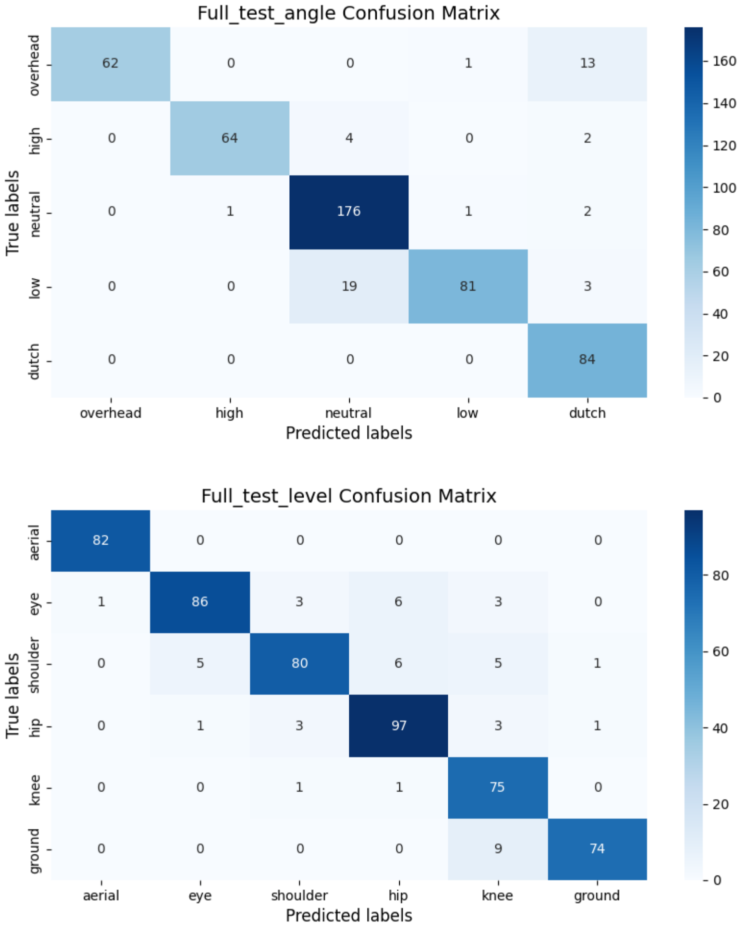
```
              precision    recall  f1-score   support

       dutch       0.86      1.00      0.93        75
        high       0.95      0.87      0.91        93
         low       0.92      0.88      0.90        98
     neutral       0.91      0.92      0.91       158
    overhead       0.95      0.94      0.95        89

    accuracy                           0.92       513
   macro avg       0.92      0.92      0.92       513
weighted avg       0.92      0.92      0.92       513
```



```
              precision    recall  f1-score   support

      aerial       0.95      0.99      0.97        87
         eye       0.93      0.77      0.85       106
      ground       0.95      0.97      0.96        79
         hip       0.85      0.85      0.85       101
        knee       0.83      0.91      0.87        70
    shoulder       0.85      0.89      0.87       100

    accuracy                           0.89       543
   macro avg       0.89      0.90      0.89       543
weighted avg       0.89      0.89      0.89       543
```

Figure B.3.4 Testing phase Classification Reports for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

Testing Phase Confusion Matrices



Full_test_angle Confusion Matrix

True labels \ Predicted labels	overhead	high	neutral	low	dutch
overhead	62	0	0	1	13
high	0	64	4	0	2
neutral	0	1	176	1	2
low	0	0	19	81	3
dutch	0	0	0	0	84

Full_test_level Confusion Matrix

True labels \ Predicted labels	aerial	eye	shoulder	hip	knee	ground
aerial	82	0	0	0	0	0
eye	1	86	3	6	3	0
shoulder	0	5	80	6	5	1
hip	0	1	3	97	3	1
knee	0	0	1	1	75	0
ground	0	0	0	0	9	74

Figure B.4.1 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=20, ILR=1e-3.

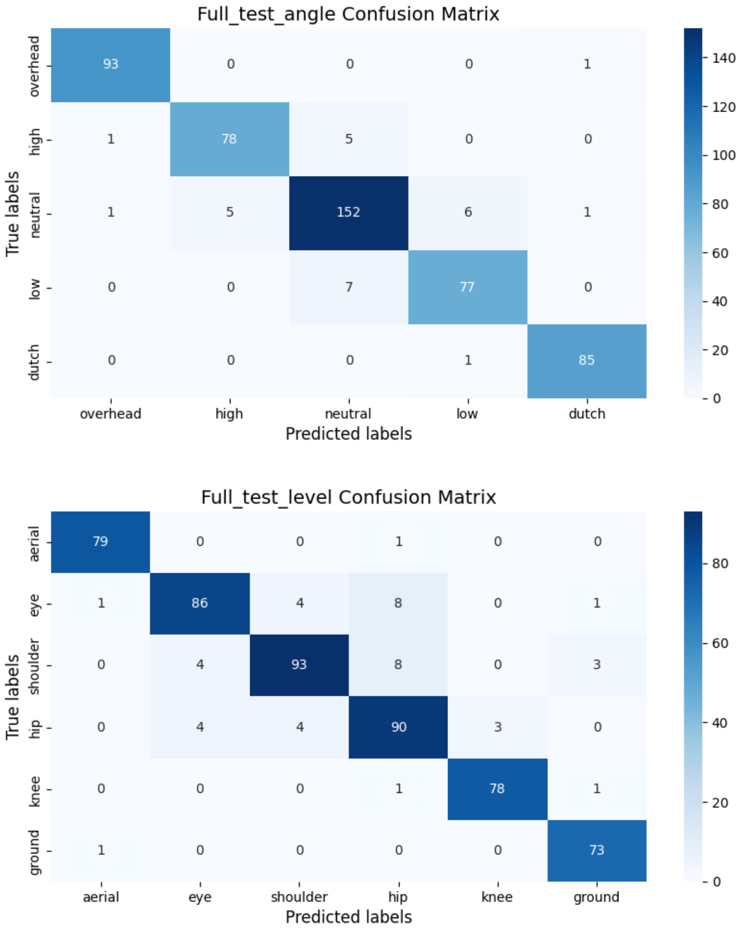


Figure B.4.2 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=30, ILR=5e-4.

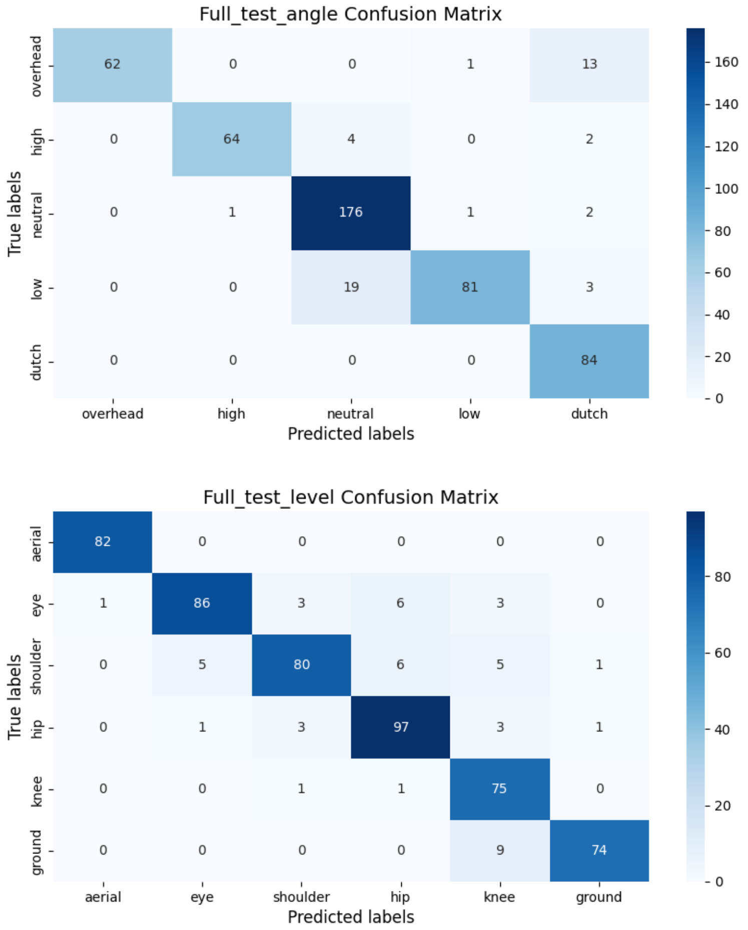


Figure B.4.3 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=40, ILR=1e-3.

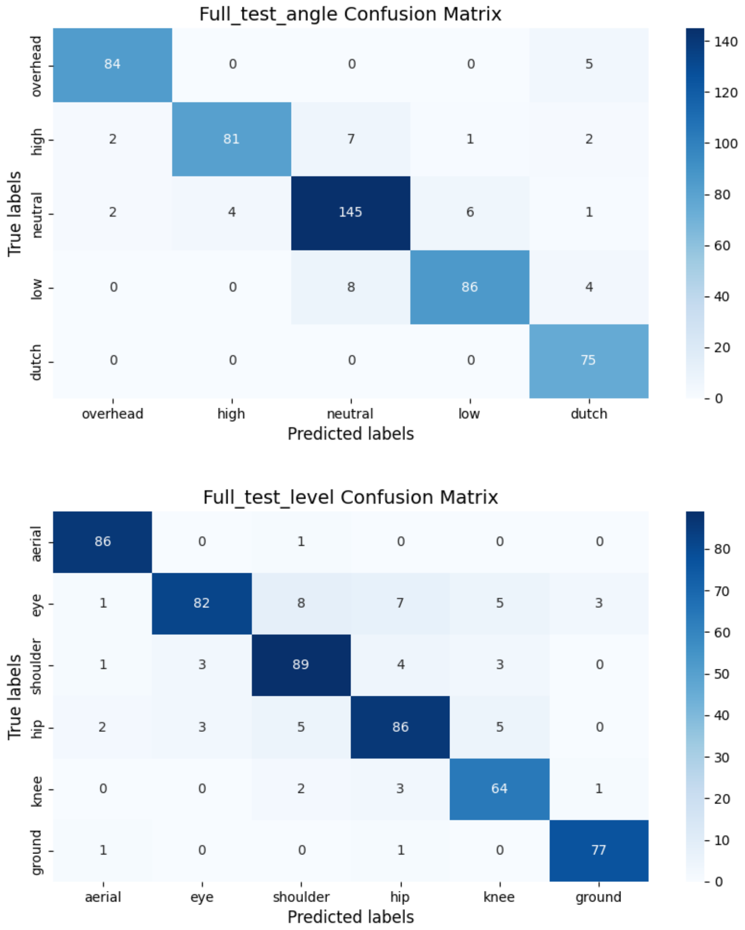
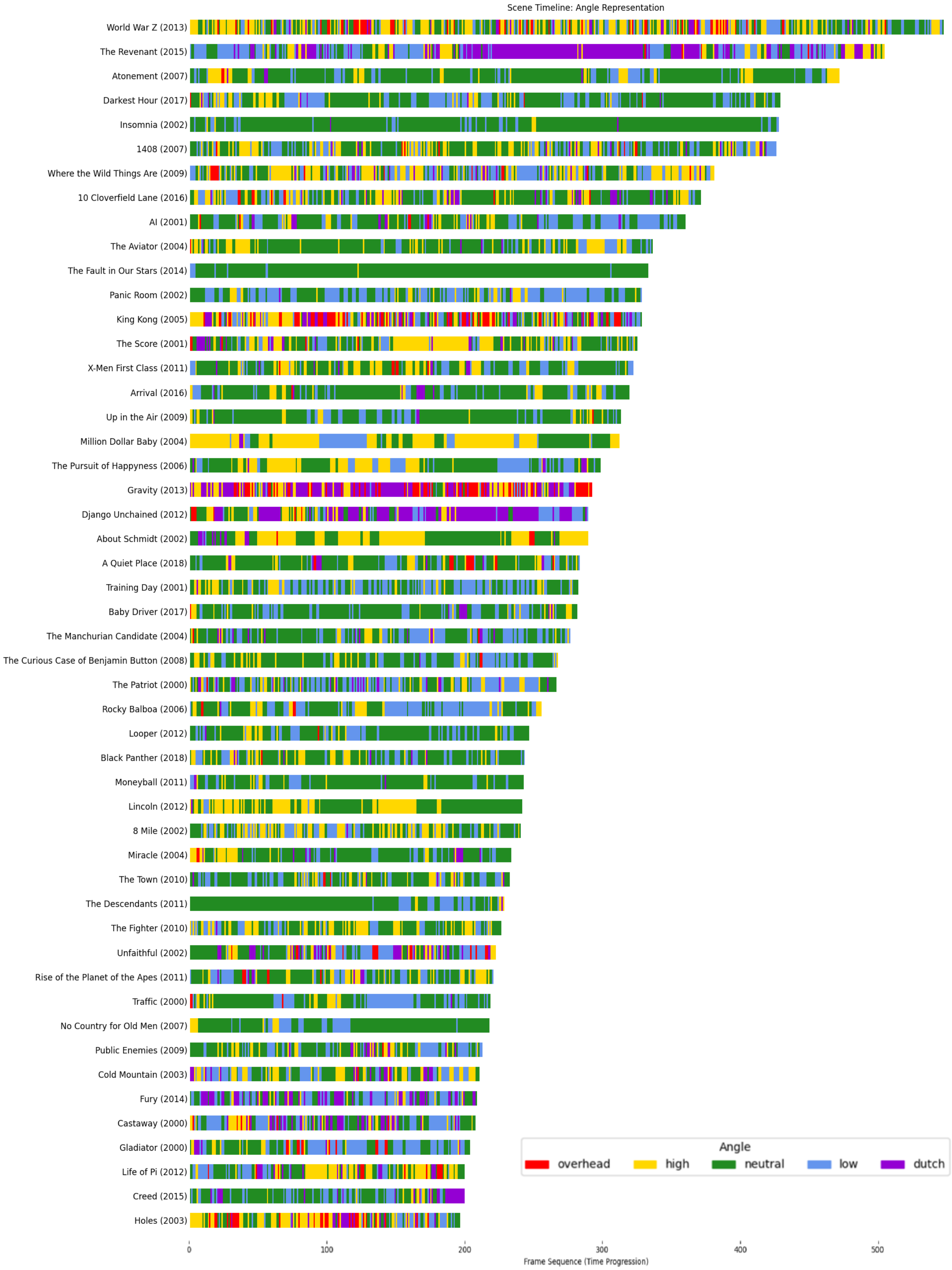


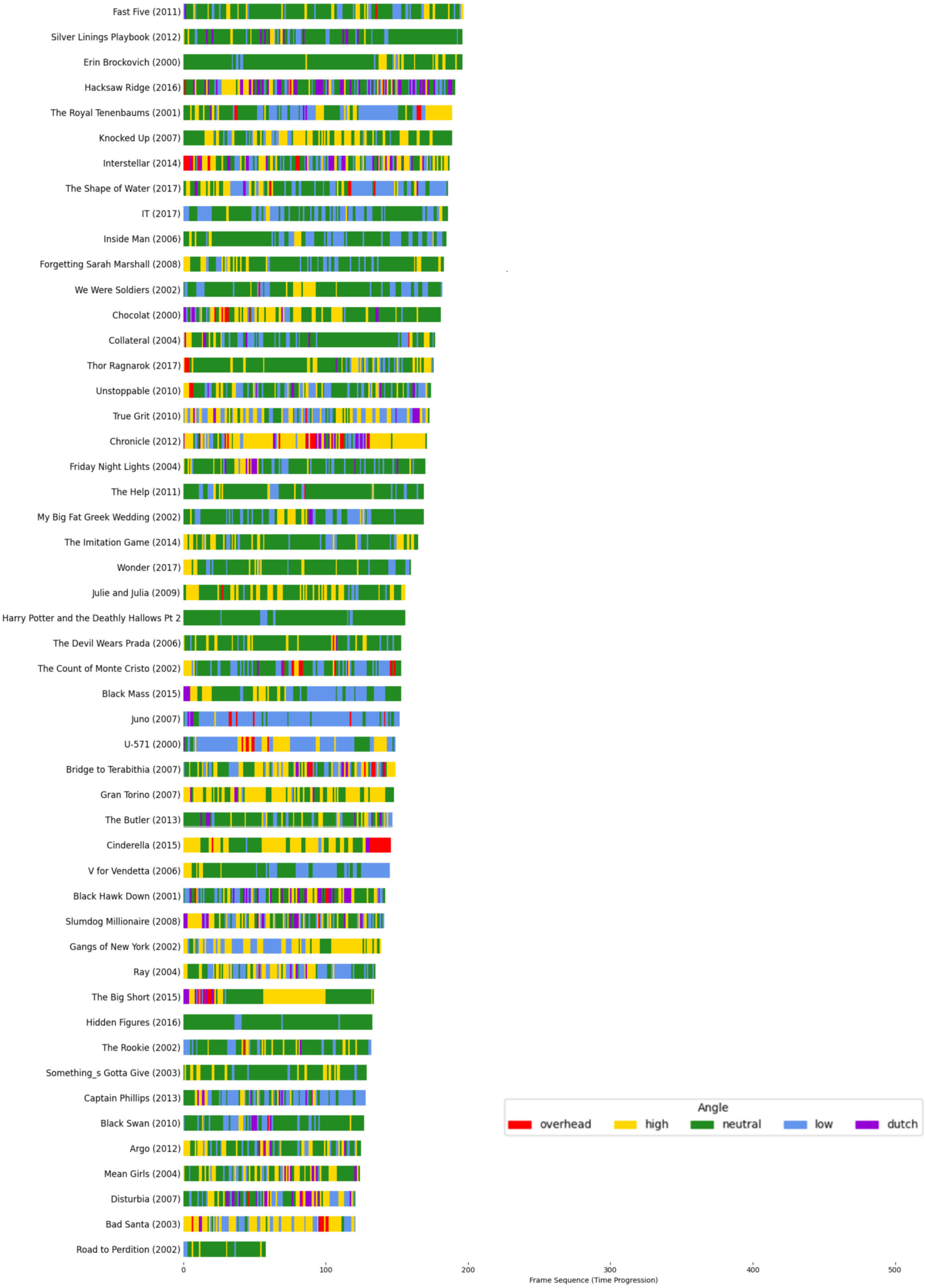
Figure B.4.4 Testing phase confusion matrices for Camera Angle and Camera Level, nEpochs=50, ILR=1e-3.

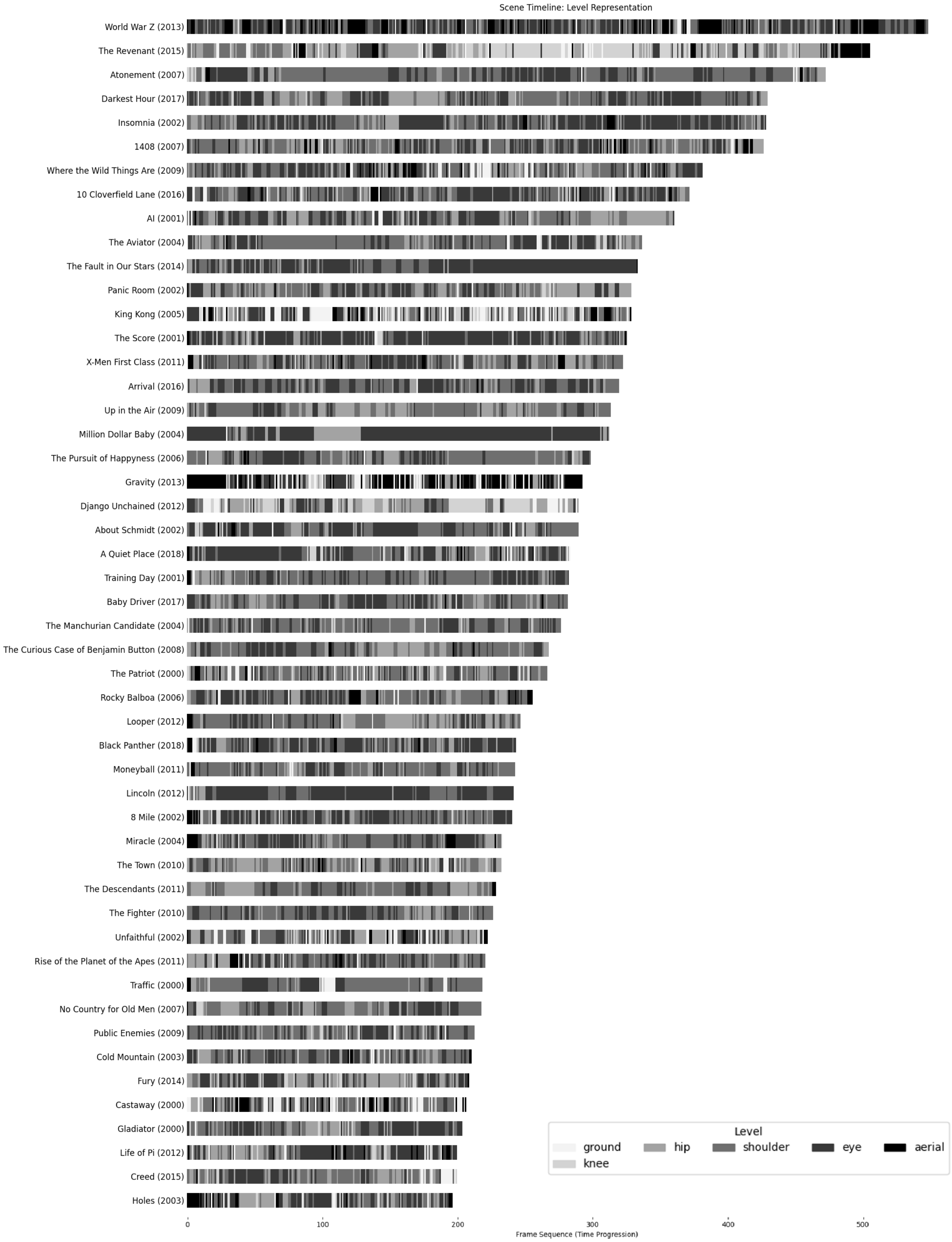
Appendix C

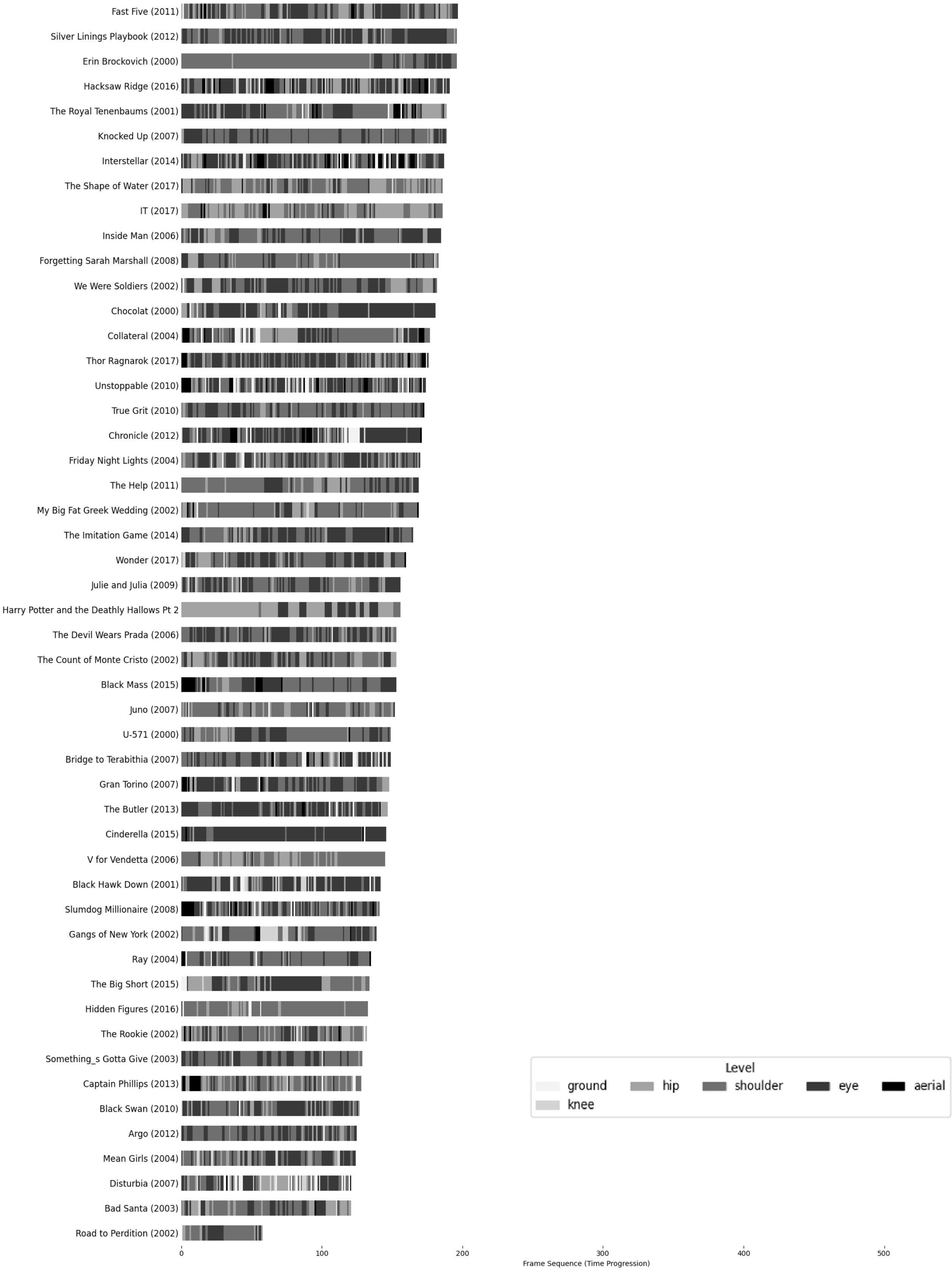
Temporal visualization of Camera Angle/Camera Level pairing

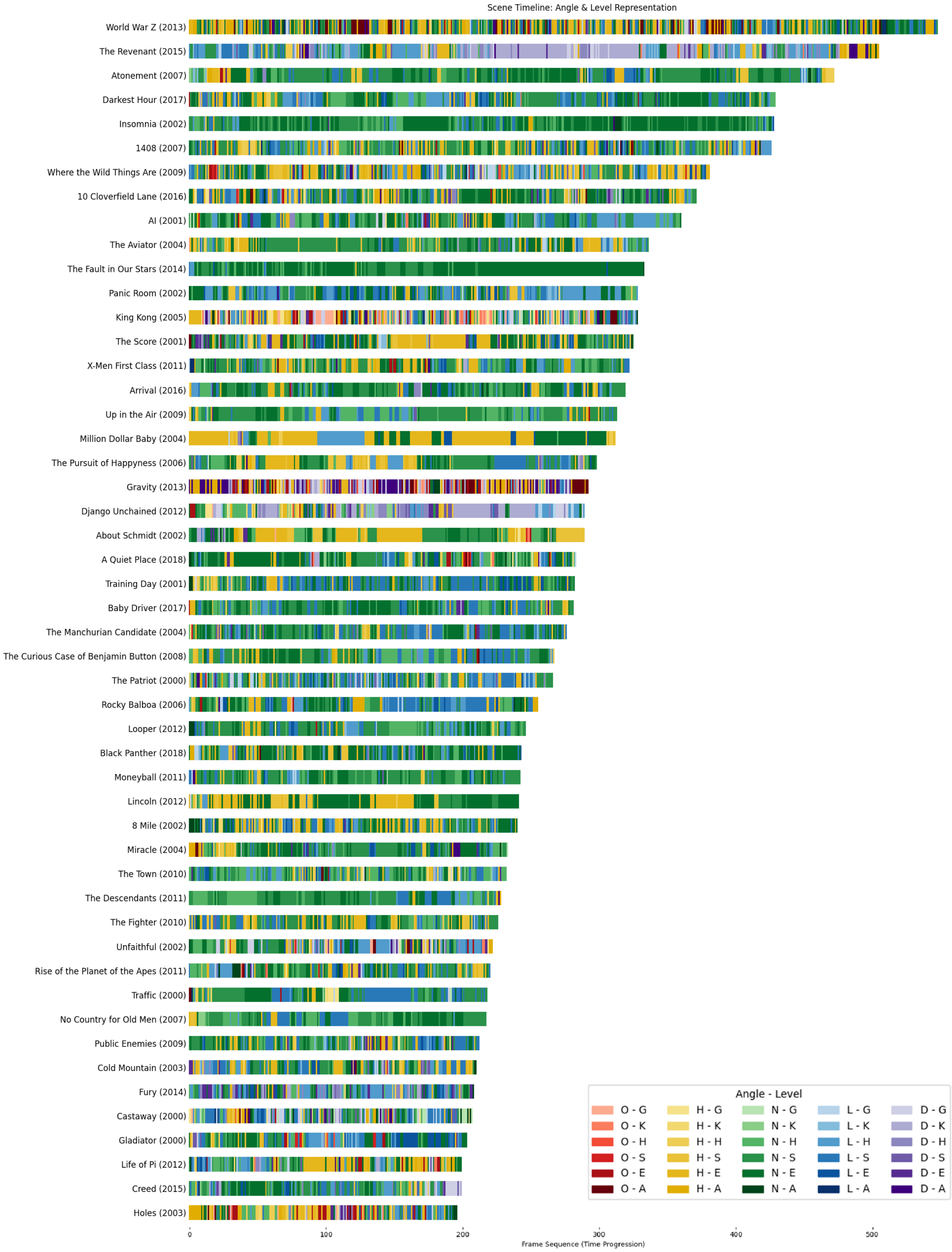
In this Appendix, the temporal visualizations of the use of Camera Angle, Camera Level and Camera Angle/Camera Level pairings are presented in full.

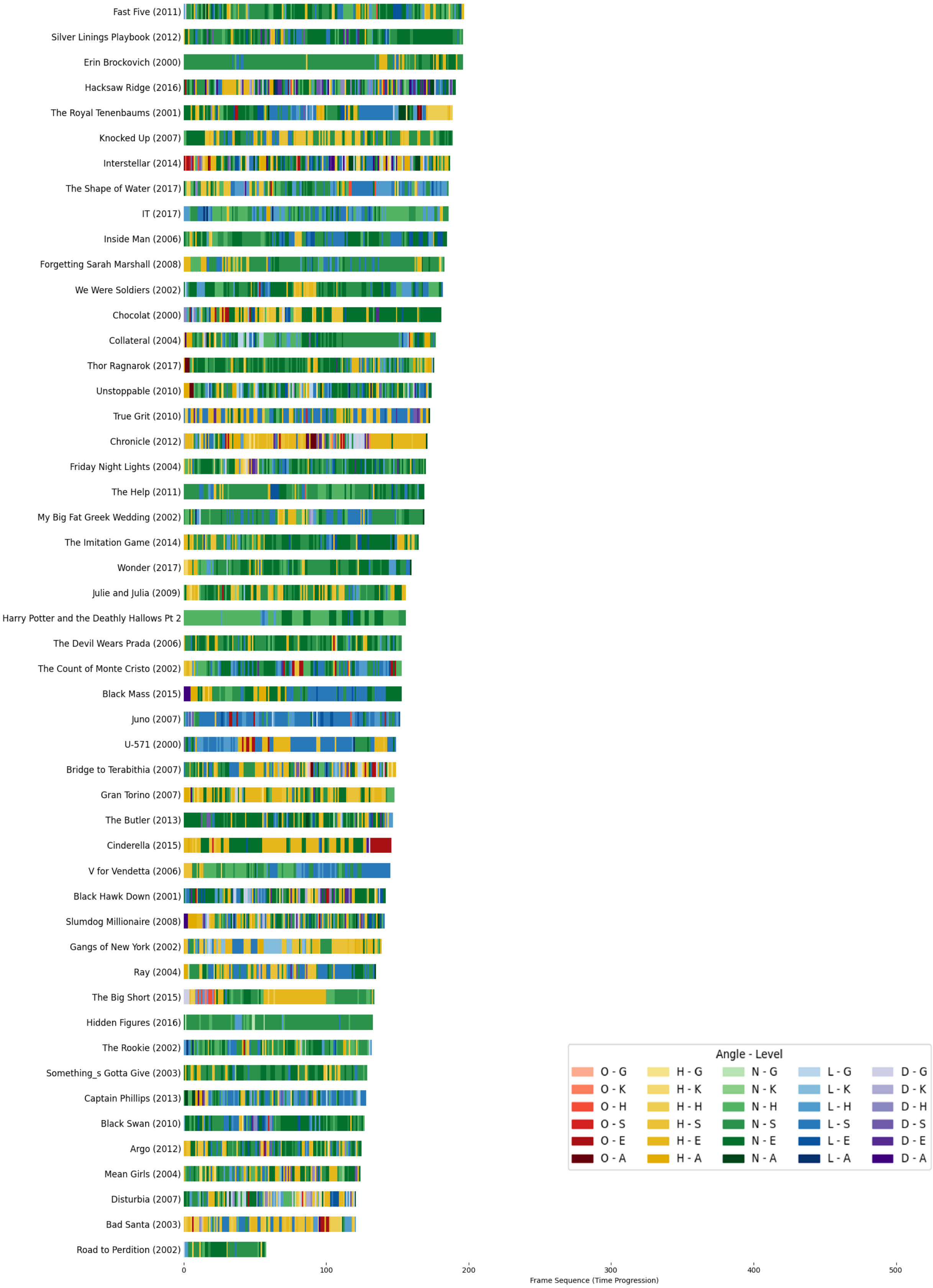

Scene Timeline: Angle Representation
World War Z (2013)
The Revenant (2015)
Atonement (2007)
Darkest Hour (2017)
Insomnia (2002)
1408 (2007)
Where the Wild Things Are (2009)
10 Cloverfield Lane (2016)
AI (2001)
The Aviator (2004)
The Fault in Our Stars (2014)
Panic Room (2002)
King Kong (2005)
The Score (2001)
X-Men First Class (2011)
Arrival (2016)
Up in the Air (2009)
Million Dollar Baby (2004)
The Pursuit of Happyness (2006)
Gravity (2013)
Django Unchained (2012)
About Schmidt (2002)
A Quiet Place (2018)
Training Day (2001)
Baby Driver (2017)
The Manchurian Candidate (2004)
The Curious Case of Benjamin Button (2008)
The Patriot (2000)
Rocky Balboa (2006)
Looper (2012)
Black Panther (2018)
Moneyball (2011)
Lincoln (2012)
8 Mile (2002)
Miracle (2004)
The Town (2010)
The Descendants (2011)
The Fighter (2010)
Unfaithful (2002)
Rise of the Planet of the Apes (2011)
Traffic (2000)
No Country for Old Men (2007)
Public Enemies (2009)
Cold Mountain (2003)
Fury (2014)
Castaway (2000)
Gladiator (2000)
Life of Pi (2012)
Creed (2015)
Holes (2003)
Angle
overhead
high
neutral
low
dutch
0
100
200
300
400
500
Frame Sequence (Time Progression)


Fast Five (2011)
Silver Linings Playbook (2012)
Erin Brockovich (2000)
Hacksaw Ridge (2016)
The Royal Tenenbaums (2001)
Knocked Up (2007)
Interstellar (2014)
The Shape of Water (2017)
IT (2017)
Inside Man (2006)
Forgetting Sarah Marshall (2008)
We Were Soldiers (2002)
Chocolat (2000)
Collateral (2004)
Thor Ragnarok (2017)
Unstoppable (2010)
True Grit (2010)
Chronicle (2012)
Friday Night Lights (2004)
The Help (2011)
My Big Fat Greek Wedding (2002)
The Imitation Game (2014)
Wonder (2017)
Julie and Julia (2009)
Harry Potter and the Deathly Hallows Pt 2
The Devil Wears Prada (2006)
The Count of Monte Cristo (2002)
Black Mass (2015)
Juno (2007)
U-571 (2000)
Bridge to Terabithia (2007)
Gran Torino (2007)
The Butler (2013)
Cinderella (2015)
V for Vendetta (2006)
Black Hawk Down (2001)
Slumdog Millionaire (2008)
Gangs of New York (2002)
Ray (2004)
The Big Short (2015)
Hidden Figures (2016)
The Rookie (2002)
Something_s Gotta Give (2003)
Captain Phillips (2013)
Black Swan (2010)
Argo (2012)
Mean Girls (2004)
Disturbia (2007)
Bad Santa (2003)
Road to Perdition (2002)
0
100
200
300
400
500
Frame Sequence (Time Progression)
Angle
overhead
high
neutral
low
dutch


Scene Timeline: Level Representation
World War Z (2013)
The Revenant (2015)
Atonement (2007)
Darkest Hour (2017)
Insomnia (2002)
1408 (2007)
Where the Wild Things Are (2009)
10 Cloverfield Lane (2016)
AI (2001)
The Aviator (2004)
The Fault in Our Stars (2014)
Panic Room (2002)
King Kong (2005)
The Score (2001)
X-Men First Class (2011)
Arrival (2016)
Up in the Air (2009)
Million Dollar Baby (2004)
The Pursuit of Happyness (2006)
Gravity (2013)
Django Unchained (2012)
About Schmidt (2002)
A Quiet Place (2018)
Training Day (2001)
Baby Driver (2017)
The Manchurian Candidate (2004)
The Curious Case of Benjamin Button (2008)
The Patriot (2000)
Rocky Balboa (2006)
Looper (2012)
Black Panther (2018)
Moneyball (2011)
Lincoln (2012)
8 Mile (2002)
Miracle (2004)
The Town (2010)
The Descendants (2011)
The Fighter (2010)
Unfaithful (2002)
Rise of the Planet of the Apes (2011)
Traffic (2000)
No Country for Old Men (2007)
Public Enemies (2009)
Cold Mountain (2003)
Fury (2014)
Castaway (2000)
Gladiator (2000)
Life of Pi (2012)
Creed (2015)
Holes (2003)
Level
ground
knee
hip
shoulder
eye
aerial
0
100
200
300
400
500
Frame Sequence (Time Progression)


Fast Five (2011)
Silver Linings Playbook (2012)
Erin Brockovich (2000)
Hacksaw Ridge (2016)
The Royal Tenenbaums (2001)
Knocked Up (2007)
Interstellar (2014)
The Shape of Water (2017)
IT (2017)
Inside Man (2006)
Forgetting Sarah Marshall (2008)
We Were Soldiers (2002)
Chocolat (2000)
Collateral (2004)
Thor Ragnarok (2017)
Unstoppable (2010)
True Grit (2010)
Chronicle (2012)
Friday Night Lights (2004)
The Help (2011)
My Big Fat Greek Wedding (2002)
The Imitation Game (2014)
Wonder (2017)
Julie and Julia (2009)
Harry Potter and the Deathly Hallows Pt 2
The Devil Wears Prada (2006)
The Count of Monte Cristo (2002)
Black Mass (2015)
Juno (2007)
U-571 (2000)
Bridge to Terabithia (2007)
Gran Torino (2007)
The Butler (2013)
Cinderella (2015)
V for Vendetta (2006)
Black Hawk Down (2001)
Slumdog Millionaire (2008)
Gangs of New York (2002)
Ray (2004)
The Big Short (2015)
Hidden Figures (2016)
The Rookie (2002)
Something_s Gotta Give (2003)
Captain Phillips (2013)
Black Swan (2010)
Argo (2012)
Mean Girls (2004)
Disturbia (2007)
Bad Santa (2003)
Road to Perdition (2002)
0
100
200
300
400
500
Frame Sequence (Time Progression)
Level
ground
hip
shoulder
eye
aerial
knee


Scene Timeline: Angle & Level Representation
World War Z (2013)
The Revenant (2015)
Atonement (2007)
Darkest Hour (2017)
Insomnia (2002)
1408 (2007)
Where the Wild Things Are (2009)
10 Cloverfield Lane (2016)
AI (2001)
The Aviator (2004)
The Fault in Our Stars (2014)
Panic Room (2002)
King Kong (2005)
The Score (2001)
X-Men First Class (2011)
Arrival (2016)
Up in the Air (2009)
Million Dollar Baby (2004)
The Pursuit of Happyness (2006)
Gravity (2013)
Django Unchained (2012)
About Schmidt (2002)
A Quiet Place (2018)
Training Day (2001)
Baby Driver (2017)
The Manchurian Candidate (2004)
The Curious Case of Benjamin Button (2008)
The Patriot (2000)
Rocky Balboa (2006)
Looper (2012)
Black Panther (2018)
Moneyball (2011)
Lincoln (2012)
8 Mile (2002)
Miracle (2004)
The Town (2010)
The Descendants (2011)
The Fighter (2010)
Unfaithful (2002)
Rise of the Planet of the Apes (2011)
Traffic (2000)
No Country for Old Men (2007)
Public Enemies (2009)
Cold Mountain (2003)
Fury (2014)
Castaway (2000)
Gladiator (2000)
Life of Pi (2012)
Creed (2015)
Holes (2003)
0
100
200
300
400
500
Frame Sequence (Time Progression)
Angle - Level
O - G
O - K
O - H
O - S
O - E
O - A
H - G
H - K
H - H
H - S
H - E
H - A
N - G
N - K
N - H
N - S
N - E
N - A
L - G
L - K
L - H
L - S
L - E
L - A
D - G
D - K
D - H
D - S
D - E
D - A


Fast Five (2011)
Silver Linings Playbook (2012)
Erin Brockovich (2000)
Hacksaw Ridge (2016)
The Royal Tenenbaums (2001)
Knocked Up (2007)
Interstellar (2014)
The Shape of Water (2017)
IT (2017)
Inside Man (2006)
Forgetting Sarah Marshall (2008)
We Were Soldiers (2002)
Chocolat (2000)
Collateral (2004)
Thor Ragnarok (2017)
Unstoppable (2010)
True Grit (2010)
Chronicle (2012)
Friday Night Lights (2004)
The Help (2011)
My Big Fat Greek Wedding (2002)
The Imitation Game (2014)
Wonder (2017)
Julie and Julia (2009)
Harry Potter and the Deathly Hallows Pt 2
The Devil Wears Prada (2006)
The Count of Monte Cristo (2002)
Black Mass (2015)
Juno (2007)
U-571 (2000)
Bridge to Terabithia (2007)
Gran Torino (2007)
The Butler (2013)
Cinderella (2015)
V for Vendetta (2006)
Black Hawk Down (2001)
Slumdog Millionaire (2008)
Gangs of New York (2002)
Ray (2004)
The Big Short (2015)
Hidden Figures (2016)
The Rookie (2002)
Something_s Gotta Give (2003)
Captain Phillips (2013)
Black Swan (2010)
Argo (2012)
Mean Girls (2004)
Disturbia (2007)
Bad Santa (2003)
Road to Perdition (2002)
0
100
200
300
400
500
Frame Sequence (Time Progression)
Angle - Level
O - G
O - K
O - H
O - S
O - E
O - A
H - G
H - K
H - H
H - S
H - E
H - A
N - G
N - K
N - H
N - S
N - E
N - A
L - G
L - K
L - H
L - S
L - E
L - A
D - G
D - K
D - H
D - S
D - E
D - A

References

[1] “Forecast 2025: Gower street announces $33 billion early global box office estimate.” [Accessed 27-02-2025]. (2024), [Online]. Available: `https://gower.st/articles/forecast-2025-gower-street-announces-33-billion-early-global-box-office-estimate/`.

[2] B. M. Cuff, S. J. Brown, L. Taylor, and D. J. Howat, “Empathy: A review of the concept,” *Emotion Review*, vol. 8, no. 2, pp. 144–153, 2016. DOI: `10.1177/1754073914558466`. eprint: `https://doi.org/10.1177/1754073914558466`. [Online]. Available: `https://doi.org/10.1177/1754073914558466`.

[3] J. Decety and P. Jackson, “A social-neuroscience perspective on empathy,” *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI*, vol. 15, Apr. 2006. DOI: `10.1111/j.0963-7214.2006.00406.x`.

[4] L. Cosmides and J. Tooby, “Cognitive adaptations for social exchange,” *Evolutionary Psychology and the Generation of Culture*, Jan. 1992.

[5] E. Tan, “Entertainment is emotion: The functional architecture of the entertainment experience,” *Media Psychology - MEDIA PSYCHOL*, vol. 11, pp. 28–51, Mar. 2008. DOI: `10.1080/15213260701853161`.

[6] A. Bartsch, “Emotional gratification in entertainment experience. why viewers of movies and television series find it rewarding to experience emotions,” *Media Psychology*, vol. 15, Jul. 2012. DOI: `10.1080/15213269.2012.693811`.

[7] M. B. Oliver and A. Raney, “Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption,” *Journal of Communication*, vol. 61, pp. 984–1004, Oct. 2011. DOI: `10.1111/j.1460-2466.2011.01585.x`.

[8] T. Grodal, *Moving Pictures: A New Theory of Film Genres, Feelings, and Cognition.* Oxford University Press, Feb. 1999, ISBN: 9780198159834. DOI: `10.1093/acprof:oso/9780198159834.001.0001`. [Online]. Available: `https://doi.org/10.1093/acprof:oso/9780198159834.001.0001`.

[9] E. Kneepkens and R. A. Zwaan, “Emotions and literary text comprehension,” *Poetics*, vol. 23, no. 1, pp. 125–138, 1995, Emotions and Cultural Products, ISSN: 0304-422X. DOI: `https://doi.org/10.1016/0304-422X(94)00021-W`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/0304422X9400021W`.

[10] L. Shen, “Mitigating psychological reactance: The role of message-induced empathy in persuasion,” *Human Communication Research*, vol. 36, no. 3, pp. 397–422, 2010. DOI: `https://doi.org/10.1111/j.1468-2958.2010.01381.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.2010.01381.x`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.2010.01381.x`.

[11] M. Green and T. Brock, “The role of transportation in the persuasiveness of public narrative,” *Journal of personality and social psychology*, vol. 79, pp. 701–21, Nov. 2000. DOI: `10.1037/0022-3514.79.5.701`.

[12] P. Taberham, *Lessons in Perception: The Avant-Garde Filmmaker as Practical Psychologist*. Berghahn Books, 2018, ISBN: 9781785336416. [Online]. Available: `http://www.jstor.org/stable/j.ctv3znzvc` (visited on 02/22/2025).

[13] H. Zettl, *Sight, Sound, Motion: Applied Media Aesthetics*. Cengage Learning, 2010, ISBN: 9781111788735. [Online]. Available: `https://books.google.sm/books?id=OfwFAAAAQBAJ`.

[14] T. Lankhuizen, K. E. Bálint, M. Savardi, E. A. Konijn, A. Bartsch, and S. Benini, “Shaping film: A quantitative formal analysis of contemporary empathy-eliciting hollywood cinema.,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 16, no. 4, pp. 704–718, Nov. 2022, ISSN: 1931-3896. DOI: `10.1037/aca0000356`. [Online]. Available: `http://dx.doi.org/10.1037/aca0000356`.

[15] S. Benini, M. Savardi, K. Bálint, A. B. Kovács, and A. Signoroni, “On the influence of shot scale on film mood and narrative engagement in film viewers,” *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 592–603, 2022. DOI: `10.1109/TAFFC.2019.2939251`.

[16] K. Balint, T. Klausch, and T. Polya, “Watching closely: Shot scale influences theory of mind response in visual narratives,” *Journal of Media Psychology*, vol. 30, pp. 1–10, Aug. 2016. DOI: `10.1027/1864-1105/a000189`.

[17] H. L. Wang and L.-F. Cheong, “Affective understanding in film,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 689–704, 2006. DOI: `10.1109/TCSVT.2006.873781`.

[18] T. Smith and J. Cutting, “A window on reality: Perceiving edited moving images,” *Current Directions in Psychological Science*, vol. 21, pp. 101–106, Mar. 2012. DOI: `10.1177/0963721412437407`.

[19] S. GEIGER and B. REEVES, “The effects of scene changes and semantic relatedness on attention to television,” *Communication Research*, vol. 20, no. 2, pp. 155–175, Apr. 1993, ISSN: 1552-3810. DOI: `10.1177/009365093020002001`. [Online]. Available: `http://dx.doi.org/10.1177/009365093020002001`.

[20] B. H. Detenber and A. Lang, “The influence of form and presentation attributes of media on emotion,” in *The Routledge Handbook of Emotions and Mass Media.* Routledge, 2011, ISBN: 9780415481601. DOI: `10.4324/9780203885390.ch16`. [Online]. Available: `http://dx.doi.org/10.4324/9780203885390.ch16`.

[21] D. Crewe, *History of the ’dutch angle’ and why movies tilt the camera*, `https://petapixel.com/2021/11/15/history-of-the-dutch-angle-and-why-movies-tilt-the-camera/`, [Accessed 26-02-2025], 2021.

[22] C. Bowen and R. Thompson, *Grammar of the Shot.* Focal Press, 2013, ISBN: 9780240526010. [Online]. Available: `https://books.google.it/books?id=Rl7IyLPt5ucC`.

[23] W. Huang and G. Olson, “Camera angle affects dominance in video-mediated communication,” Apr. 2002, pp. 716–717. DOI: `10.1145/506443.506562`.

[24] R. N. Kraft, “The influence of camera angle on comprehension and retention of pictorial events,” *Memory & Cognition*, vol. 15, no. 4, pp. 291–307, Jul. 1987, ISSN: 1532-5946. DOI: `10.3758/bf03197032`. [Online]. Available: `http://dx.doi.org/10.3758/BF03197032`.

[25] Y. Tsivian, *Cinemetrics - movie measurement and study tool database*, `https://cinemetrics.uchicago.edu/`, [Accessed 27-02-2025], 2005.

[26] T. H. Trojahn and R. Goularte, “Temporal video scene segmentation using deep-learning,” *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 17 487–17 513, Feb. 2021, ISSN: 1573-7721. DOI: `10.1007/s11042-020-10450-2`. [Online]. Available: `http://dx.doi.org/10.1007/s11042-020-10450-2`.

[27] G. Touros and T. Giannakopoulos, “Video soundtrack evaluation with machine learning: Data availability, feature extraction, and classification,” in *Advances in Speech and Music Technology.* Springer International Publishing, Sep. 2022, pp. 137–157, ISBN: 9783031184444. DOI: `10.1007/978-3-031-18444-4_7`. [Online]. Available: `http://dx.doi.org/10.1007/978-3-031-18444-4_7`.

[28] K. Kleist, *Machine learning & cinema: Color grading style classification through vectorscopes and cnns*, `https://webthesis.biblio.polito.it/29472/`, [Accessed 27-02-2025], 2023.

[29] M. Savardi, A. Signoroni, P. Migliorati, and S. Benini, "Shot scale analysis in movies by convolutional neural networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2620–2624. DOI: `10.1109/ICIP.2018.8451474`.

[30] M. Savardi, A. B. Kovács, A. Signoroni, and S. Benini, "Recognition of camera angle and camera level in movies from single frames," in *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences Workshops*, ser. IMXw '23, Nantes, France: Association for Computing Machinery, 2023, pp. 79–85, ISBN: 9798400708459. DOI: `10.1145/3604321.3604334`. [Online]. Available: `https://doi.org/10.1145/3604321.3604334`.

[31] K. Thompson, *Mapping contemporary cinema - a short guide to cinemetrics*, `https://mcc.sllf.qmul.ac.uk/?p=1859`, [Accessed 01-03-2025], 2017.

[32] K. Viarshynina, *Cinemetrics: Creative ways to measure and visualize movie data*, `https://mastersofmedia.hum.uva.nl/2014/04/cinemetrics-creative-ways-to-measure-and-visualize-movie-data/`, [Accessed 01-03-2025], 2014.

[33] A. F. C. e Silva, "Cinemetrics: Creative approaches to visualize and analyze movie data," M.S. thesis, Instituto Superior de Engenharia do Porto, Nov. 2021.

[34] F. Brodbeck, *Cinemetrics*, `https://cinemetrics.site/`, [Accessed 01-03-2025], 2011.

[35] B. Dawes, *Cinema redux*, `https://brendandawes.com/projects/cinemaredux`, [Accessed 01-03-2025], 2004.

[36] *Moviebarcode*, `https://moviebarcode.tumblr.com/`, [Accessed 01-03-2025].

[37] C. Clark, *The colors of motion*, `https://thecolorsofmotion.com/`, [Accessed 01-03-2025], 2015.

[38] J. Meyers-Levy and L. A. Peracchio, "Getting an angle in advertising: The effect of camera angle on product evaluations," *Journal of Marketing Research*, vol. 29, no. 4, p. 454, Nov. 1992, ISSN: 0022-2437. DOI: `10.2307/3172711`. [Online]. Available: `http://dx.doi.org/10.2307/3172711`.

[39] I. Bevers, *A change of perspecitive : The influence of camera angles on product evaluation*, Aug. 2007. [Online]. Available: `http://essay.utwente.nl/58759/`.

[40] M. Merkt, A.-L. Weingärtner, and S. Schwan, "Digital images are hard to resist: Teaching viewers about the effects of camera angle does not reduce the camera angle's impact on power judgments," *Acta Psychologica*, vol. 229, p. 103 687, Sep. 2022, ISSN: 0001-6918. DOI: `10.1016/j.actpsy.2022.103687`. [Online]. Available: `http://dx.doi.org/10.1016/j.actpsy.2022.103687`.

[41] A. Rao, J. Wang, L. Xu, *et al.*, *A unified framework for shot type classification based on subject centric lens*, 2020. DOI: `10.48550/ARXIV.2008.03548`. [Online]. Available: `https://arxiv.org/abs/2008.03548`.

[42] K. Ouenniche, R. Tapu, and T. Zaharia, "A deep learning-based approach for camera motion classification," in *2021 9th European Workshop on Visual Information Processing (EUVIP)*, IEEE, Jun. 2021, pp. 1–6. DOI: `10.1109/euvip50544.2021.9483961`. [Online]. Available: `http://dx.doi.org/10.1109/EUVIP50544.2021.9483961`.

[43] S. Benini, M. Svanera, N. Adami, R. Leonardi, and A. B. Kovács, "Shot scale distribution in art films," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16 499–16 527, Feb. 2016, ISSN: 1573-7721. DOI: `10.1007/s11042-016-3339-9`. [Online]. Available: `http://dx.doi.org/10.1007/s11042-016-3339-9`.

[44] K. Bálint, T. Klausch, and T. Pólya, "Watching closely: Shot scale influences theory of mind response in visual narratives," *Journal of Media Psychology*, vol. 30, no. 3, pp. 150–159, Jul. 2018, ISSN: 2151-2388. DOI: `10.1027/1864-1105/a000189`. [Online]. Available: `http://dx.doi.org/10.1027/1864-1105/a000189`.

[45] M. Savardi, S. Benini, and A. B. Kovács, *Recognition of camera angle and camera level in movies from single frames*, `https://cinescale.github.io/camera_al/`, [Accessed 27-02-2025], 2023.

[46] F. S. Rejects, *One perfect shot database | perfect movie frames from film school rejects*, `https://shots.filmschoolrejects.com/`, [Accessed 27-02-2025].

[47] A. B. Jung, *imgaug*, `https://github.com/aleju/imgaug`, [Online; accessed 30-Oct-2018], 2018.

[48] A. Paszke, S. Gross, F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[49] M. L. McHugh, "The chi-square test of independence," *Biochemia Medica*, pp. 143–149, 2013, ISSN: 1846-7482. DOI: `10.11613/bm.2013.018`. [Online]. Available: `http://dx.doi.org/10.11613/BM.2013.018`.

[50] H. Cramér, *Mathematical Methods of Statistics* (Princeton Mathematical Series). Princeton University Press, 1946, vol. 9.

[51] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20 150 202, Apr. 2016, ISSN: 1471-2962. DOI: `10.1098/rsta.2015.0202`. [Online]. Available: `http://dx.doi.org/10.1098/rsta.2015.0202`.